

Dear Examiner,

The attached search was run with the most recently released version of Compugen's search software, GenCore 5. With this update, several changes have occurred in the results of FrameSearches (protein query sequence vs nucleic acid databases or nucleic acid query sequence vs protein databases).

In reference to FrameSearches:

- The output format has been improved so that it more closely resembles the format for standard search output.
- Calculation of Percent Similarity has been changed for FrameSearches. The new method of calculation is more similar to the method used in NCBI's BLAST algorithm. The same results are found in the same order using GenCore 5 and the previous version of GenCore, but Percent Similarities are lower in GenCore 5 results.

- The formula for % similarity calculation is:

$$100 * \frac{\text{matches} + \text{conservative substitutions}}{\text{alignment_length}}$$

where "matches" is the number of identical matches and "conservative substitutions" is the number of non-identical positive matches.

- GenCore 4.5 considers the match Thr vs GCT (Ala) to be a similarity since BLOSUM62 gives score of 0 to this match. It is marked by ':::' in the alignment:

```
Qy 46 AspSerThrAspAla.Met..Gly 52
      |||||::: ||| ::: |||
Db 605 GATTCCGCTGCTGCTAATTTGGC 628
```

GenCore 5 requires a positive score to consider a non-identical match a similarity, therefore the same 'match' is not emphasized in the new alignment:

```
Qy 46 AspSerThrAspAla.Met..Gly 52
      ||||| ||| ::: |||
Db 605 GATTCCGCTGCTGCTAATTTGGC 628
```

If you have any questions, please feel free to contact one of the searchers in Biotech/Chem Library.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 10:34:25 ; Search time 66 seconds

(without alignments)
325.051 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186

Perfect score: 848
Sequence: 1 PVTNVCQNLVWDLSTQIFCH.....DVTVLDPDYRGVPIPLFY 161

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_101002:.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	279	23	AAE18419
2	845	99.6	279	23	AAE18426
3	838	98.8	279	23	AAE18435
4	837	98.7	279	23	AAE18436
5	834	98.3	279	23	AAE18418
6	834	98.3	279	23	AAE18422
7	834	98.3	279	23	AAE18424
8	834	98.3	279	23	AAE18433
9	834	98.3	300	16	AAE18433
10	834	98.3	300	16	AAE18433

11	834	98.3	300	21	AAE16009	E. coli proliferat
12	834	98.3	300	21	AAE16009	E. coli PC31 FlmH
13	834	98.3	300	21	AAE16009	Escherichia coli t
14	833	98.2	279	23	AAE18427	Escherichia coli s
15	832	98.1	279	23	AAE18417	Escherichia coli s
16	831	98.0	279	23	AAE18413	Escherichia coli s
17	831	98.0	279	23	AAE18432	Escherichia coli s
18	830	97.9	300	16	AAE18432	FlmH protein deriv
19	829	97.8	279	22	AAE18420	Adhesin protein, F
20	829	97.8	279	22	AAE18420	Escherichia coli F
21	829	97.8	279	23	AAE18428	Escherichia coli s
22	829	97.8	279	23	AAE18430	Escherichia coli s
23	829	97.8	279	23	AAE18415	Immunogenic comple
24	829	97.8	304	22	AAE18416	FlmH protein deriv
25	828	97.6	300	16	AAE18429	Escherichia coli s
26	827	97.5	279	23	AAE18429	Escherichia coli s
27	827	97.5	280	23	AAE18420	Escherichia coli s
28	826	97.4	279	23	AAE18415	Escherichia coli s
29	826	97.4	279	23	AAE18416	Escherichia coli s
30	826	97.4	279	23	AAE18434	Escherichia coli s
31	825	97.3	279	23	AAE18421	Escherichia coli s
32	825	97.3	300	16	AAE18414	FlmH protein deriv
33	825	97.3	300	16	AAE18423	FlmH protein deriv
34	825	97.3	300	16	AAE18415	FlmH protein deriv
35	824	97.2	300	16	AAE18425	FlmH protein deriv
36	823	97.1	279	23	AAE18425	Escherichia coli s
37	822	96.9	279	23	AAE18414	FlmH protein deriv
38	819	96.6	279	23	AAE18414	Escherichia coli s
39	819	96.6	279	23	AAE18414	Escherichia coli s
40	819	96.6	279	23	AAE18431	Escherichia coli s
41	818	96.5	300	16	AAE18431	FlmH protein deriv
42	816	96.2	408	23	AAE18431	Fusion protein prt
43	815	96.1	300	16	AAE18431	FlmH protein deriv
44	814	96.0	300	16	AAE18431	FlmH protein deriv
45	814	96.0	300	16	AAE18431	FlmH protein deriv

ALIGNMENTS

RESULT 1	
AAE18419	
ID	AAE18419 standard; Protein: 279 AA.
XX	
AC	AAE18419;
XX	
DT	07-MAY-2002 (first entry)
XX	
DE	Escherichia coli strain B240 FlmH protein.
XX	
KW	FlmH; immune response; antibacterial; enterobacillus-related disease;
KW	therapy; vaccine; urinary tract infection; bladder.
XX	
OS	Escherichia coli B240.
XX	
FT	Key
FT	Misc-difference 201
FT	Location/Qualifiers
XX	
PN	W0200204496-A2.
XX	
PD	17-JAN-2002.
XX	
PF	06-JUL-2001; 2001WO-US21525.
XX	
PR	07-JUL-2000; 2000US-216750P.
XX	
PA	(MEDI-) MEDIMUNE INC.
XX	
PI	Langemann S, Revel A, Auguste C, Burtin J;
XX	
DR	WPI: 2002-171702/22.
DR	N-PSDB; AAD29358.

XX New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
XX
XX
PS Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain B240 FimH protein.
XX
SQ Sequence 279 AA:

Query Match 100.0%; Score 848; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 2,1e-83;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PVNVNGQNLVVDLSTQICFCHNDYPTETIDYVTLQSGSAYGVLNFSGTVKXSGSSYPP 60
DB 26 PVNVNGQNLVVDLSTQICFCHNDYPTETIDYVTLQSGSAYGVLNFSGTVKXSGSSYPP 85
QY 61 TTSETPRVYVNSRTDKPMPVALYLTLPVSSAGGLVYKAGSLAVLILRQTNNNYNSDDQFV 120
DB 86 TTSETPRVYVNSRTDKPMPVALYLTLPVSSAGGLVYKAGSLAVLILRQTNNNYNSDDQFV 145
QY 121 WNIYANDVVPPTGCDVSARDVTVTLDPYRGSVPIPLTV 161
DB 166 WNIYANDVVPPTGCDVSARDVTVTLDPYRGSVPIPLTV 186

RESULT 2
AAE18426 1
ID AAE18426 standard; Protein: 279 AA.
XX
AC AAE18426;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC60 FimH protein.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli EC60.
XX
FH Key Location/Qualifiers
FT Misc-difference 27 /note= "Encoded by GCC"
FT Misc-difference 66 /note= "Encoded by AGC"
FT Misc-difference 70 /note= "Encoded by AGC"
FT Misc-difference 70 /note= "Encoded by AGT"
FT Misc-difference 78 /note= "Encoded by AAT"
FT Misc-difference 119 /note= "Encoded by GCG"
FT Misc-difference 176 /note= "Encoded by CCG"
FT Misc-difference 201 /note= "Encoded by ACC"
FT Misc-difference 269 /note= "Encoded by CAA"
FT Misc-difference 273 /note= "Encoded by GGC"
XX
XX
PN WO200204496-A2.

XX
PD 17-JAN-2002.
XX
XX 06-JUL-2001; 2001WO-US21525.
XX
XX 07-JUL-2000; 2000US-216750P.
XX
XX (MEDI-) MEDIMUNE INC.
XX
XX Langermann S, Revel A, Auguste C, Burlain J;
XX WPI; 2002-171702/22.
XX N-PSDB; AAD29365.
XX
XX
XX
XX
PS Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC60 FimH protein.
XX
SQ Sequence 279 AA:

Query Match 99.6%; Score 845; DB 23; Length 279;
Best Local Similarity 99.4%; Pred. No. 4.4e-83;
Matches 160; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PVNVNGQNLVVDLSTQICFCHNDYPTETIDYVTLQSGSAYGVLNFSGTVKXSGSSYPP 60
DB 26 PVNVNGQNLVVDLSTQICFCHNDYPTETIDYVTLQSGSAYGVLNFSGTVKXSGSSYPP 85
QY 61 TTSETPRVYVNSRTDKPMPVALYLTLPVSSAGGLVYKAGSLAVLILRQTNNNYNSDDQFV 120
DB 86 TTSETPRVYVNSRTDKPMPVALYLTLPVSSAGGLVYKAGSLAVLILRQTNNNYNSDDQFV 145
QY 121 WNIYANDVVPPTGCDVSARDVTVTLDPYRGSVPIPLTV 161
DB 146 WNIYANDVVPPTGCDVSARDVTVTLDPYRGSVPIPLTV 186

RESULT 3
AAE18435
ID AAE18435 standard; Protein: 279 AA.
XX
AC AAE18435;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain G162 FimH protein.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli G162.
XX
FH Key Location/Qualifiers
FT Misc-difference 176 /note= "Encoded by CCG"
FT Misc-difference 201 /note= "Encoded by ACC"
XX
XX
PN WO200204496-A2.

PD 17-JAN-2002.
XX
XX 06-JUL-2001: 2001WO-US21525.
PF
XX 07-JUL-2000: 2000US-216750P.
PR
XX (MEDI-) MEDIMUNE INC.
PA
XX Langermann S, Revel A, Auguste C, Burlein J;
PI
XX WPI: 2002-171702/22.
DR
XX N-PSDB: AAD29382.
DR
XX
XX New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
PT
XX
XX Claim 3: Page 88-89; 101pp: English.
XX
XX The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli F1mH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain G162 F1mH protein.
XX
XX Sequence 279 AA:
SQ
Query Match 98.8%; Score 838; DB 23; Length 279;
Best Local Similarity 98.1%; Pred. No. 2.5e-82;
Matches 158; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 PVAWVGONLVVDLSTQIFCHNDYPERITTDVYTLQRGSAVGGVLSNFGSTVYKSGSSYFP 60
DB 26 PVAWVGONLVVDLSTQIFCHNDYPERITTDVYTLQRGSAVGGVLSNFGSTVYKSGSSYFP 85
QY 61 TTSETPRVYVNSRTDKMPVALYLPVSSAGGLVYKAGSLAVLILRQTNNNYNSDDQFV 120
DB 86 TTSETPRVYVNSRTDKMPVALYLPVSSAGGLVYKAGSLAVLILRQTNNNYNSDDQFV 145
QY 121 WNIYANDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
DB 146 WNIYANDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186
RESULT 4
AAE18436
ID AAE18436 standard; Protein: 279 AA.
XX
XX AAE18436:
AC
XX 07-MAY-2002 (first entry)
DT
XX
XX Escherichia coli F1mH consensus protein.
DE
XX
XX F1mH: immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
XX Escherichia coli.
OS
XX
XX WO200204496-A2.
PN
XX
XX 17-JAN-2002.
PD
XX
XX 06-JUL-2001: 2001WO-US21525.
PF
XX
XX 07-JUL-2000: 2000US-216750P.
PR
XX
XX (MEDI-) MEDIMUNE INC.
PA

XX
XX Langermann S, Revel A, Auguste C, Burlein J;
PI
XX WPI: 2002-171702/22.
DR
XX
XX New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
PT
XX
XX Claim 3: Fig 2; 101pp: English.
XX
XX The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli F1mH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli F1mH consensus protein.
XX
XX Sequence 279 AA:
SQ
Query Match 98.7%; Score 837; DB 23; Length 279;
Best Local Similarity 98.1%; Pred. No. 3.3e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PVAWVGONLVVDLSTQIFCHNDYPERITTDVYTLQRGSAVGGVLSNFGSTVYKSGSSYFP 60
DB 26 PVAWVGONLVVDLSTQIFCHNDYPERITTDVYTLQRGSAVGGVLSNFGSTVYKSGSSYFP 85
QY 61 TTSETPRVYVNSRTDKMPVALYLPVSSAGGLVYKAGSLAVLILRQTNNNYNSDDQFV 120
DB 86 TTSETPRVYVNSRTDKMPVALYLPVSSAGGLVYKAGSLAVLILRQTNNNYNSDDQFV 145
QY 121 WNIYANDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
DB 146 WNIYANDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186
RESULT 5
AAE18418
ID AAE18418 standard; Protein: 279 AA.
XX
XX AAE18418:
AC
XX 07-MAY-2002 (first entry)
DT
XX
XX Escherichia coli strain B238 F1mH protein.
DE
XX
XX F1mH: immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
XX Escherichia coli B238.
OS
XX
XX
FH Key Location/Qualifiers
FT MISC-difference 24 /note- "Encoded by ATT"
FT MISC-difference 26 /note- "Encoded by CCC"
FT MISC-difference 176 /note- "Encoded by CCT"
FT MISC-difference 201 /note- "Encoded by ACC"
FT MISC-difference 273 /note- "Encoded by GCC"
FT MISC-difference 274 /note- "Encoded by GTG"
FT
XX
XX WO200204496-A2.
PN
XX
XX 17-JAN-2002.
PD

XX 06-JUL-2001; 2001WO-US21525.
XX
XX 07-JUL-2000; 2000US-216750P.
XX
XX (MEDI-) MEDIMUNE INC.
XX
XX Langermann S, Revel A, Auguste C, Burtlein J;
XX
XX WPI: 2002-171702/22.
XX N-PSDB; AAD29357.
XX
XX
XX New immunogenic polypeptide, useful as vaccine for protecting against
XX an enterobacillus-related disease in a patient at risk of contracting
XX such disease, e.g. urinary tract infection or a bladder infection
XX
XX Claim 3; Fig 2; 101pp; English.
XX
XX The invention relates to bacterial immunogenic agents for administration
XX to humans and non-human animals to stimulate an immune response. The
XX invention also relates to methods for vaccination of mammalian species
XX with variants of E. coli FimH protein derived from different strains of
XX E. coli. The vaccine composition or the antibody is useful for protecting
XX against and treating an enterobacillus-related disease in a patient
XX afflicted or at a risk of contracting the disease. In particular, the
XX disease is a urinary tract or bladder infection. The disease is caused
XX by a bacterium of the family Enterobacteriaceae, particularly E. coli.
XX The present sequence is Escherichia coli strain B238 FimH protein.
SQ Sequence 279 AA;
Query Match 98.3%; Score 834; DB 23; Length 279;
Best Local Similarity 99.4%; Pred. No. 6.9e-82;
Matches 158; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 3 VAVNGONLVVDLSTQIFCHNDYPTITDYVTLORGSAVGGVLSNFGSTVYSGSSYPPTT 62
DB 28 VAVNGONLVVDLSTQIFCHNDYPTITDYVTLORGSAVGGVLSNFGSTVYSGSSYPPTT 87
QY 63 SETPRVYNSRTDKPMPVALYLPVSSAGGLYKAGSLIAVLILROTNMYNSDDFOFVWN 122
DB 88 SETPRVYNSRTDKPMPVALYLPVSSAGGLYKAGSLIAVLILROTNMYNSDDFOFVWN 147
QY 123 IYANDVVPVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
DB 148 IYANDVVPVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186
RESULT 6
AAE18422
ID AAE18422 standard; Protein: 279 AA.
XX
XX AAE18422;
XX
XX 07-MAY-2002 (first entry)
XX
XX Escherichia coli strain EC42 FimH protein.
XX
XX
XX FimH; immune response; antibacterial; enterobacillus-related disease;
XX therapy; vaccine; urinary tract infection; bladder.
XX
XX Escherichia coli EC42.
XX
XX WO200204496-A2.
XX
XX
XX 17-JAN-2002.
XX
XX
XX 06-JUL-2001; 2001WO-US21525.
XX
XX 07-JUL-2000; 2000US-216750P.
XX
XX (MEDI-) MEDIMUNE INC.
XX

PI Langermann S, Revel A, Auguste C, Burtlein J;
XX
XX WPI: 2002-171702/22.
XX
XX N-PSDB; AAD29361.
XX
XX
XX New immunogenic polypeptide, useful as vaccine for protecting against
XX an enterobacillus-related disease in a patient at risk of contracting
XX such disease, e.g. urinary tract infection or a bladder infection
XX
XX Claim 3; Fig 2; 101pp; English.
XX
XX The invention relates to bacterial immunogenic agents for administration
XX to humans and non-human animals to stimulate an immune response. The
XX invention also relates to methods for vaccination of mammalian species
XX with variants of E. coli FimH protein derived from different strains of
XX E. coli. The vaccine composition or the antibody is useful for protecting
XX against and treating an enterobacillus-related disease in a patient
XX afflicted or at a risk of contracting the disease. In particular, the
XX disease is a urinary tract or bladder infection. The disease is caused
XX by a bacterium of the family Enterobacteriaceae, particularly E. coli.
XX The present sequence is Escherichia coli strain EC42 FimH protein.
SQ Sequence 279 AA;
Query Match 98.3%; Score 834; DB 23; Length 279;
Best Local Similarity 98.1%; Pred. No. 6.9e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PVVNVGONLVVDLSTQIFCHNDYPTITDYVTLORGSAVGGVLSNFGSTVYSGSSYPFP 60
DB 26 PAVNVGONLVVDLSTQIFCHNDYPTITDYVTLORGSAVGGVLSNFGSTVYSGSSYPFP 85
QY 61 TTSEPRVYNSRTDKPMPVALYLPVSSAGGLYKAGSLIAVLILROTNMYNSDDFOFV 120
DB 86 TTSEPRVYNSRTDKPMPVALYLPVSSAGGLYKAGSLIAVLILROTNMYNSDDFOFV 145
QY 121 WNIYANDVVPVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
DB 146 WNIYANDVVPVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186
RESULT 7
AAE18424
ID AAE18424 standard; Protein: 279 AA.
XX
XX AAE18424;
XX
XX 07-MAY-2002 (first entry)
XX
XX Escherichia coli strain EC56 FimH protein.
XX
XX
XX FimH; immune response; antibacterial; enterobacillus-related disease;
XX therapy; vaccine; urinary tract infection; bladder.
XX
XX Escherichia coli EC56.
XX
XX WO200204496-A2.
XX
XX
XX 17-JAN-2002.
XX
XX
XX 06-JUL-2001; 2001WO-US21525.
XX
XX 07-JUL-2000; 2000US-216750P.
XX
XX (MEDI-) MEDIMUNE INC.
XX
XX Langermann S, Revel A, Auguste C, Burtlein J;
XX
XX WPI: 2002-171702/22.
XX
XX N-PSDB; AAD29363.
XX
XX
XX New immunogenic polypeptide, useful as vaccine for protecting against
XX an enterobacillus-related disease in a patient at risk of contracting

PT such disease, e.g. urinary tract infection or a bladder infection -
 XX
 XX Claim 3; Fig 2; 101pp; English.
 XX
 CC The invention relates to bacterial immunogenic agents for administration
 CC to humans and non-human animals to stimulate an immune response. The
 CC invention also relates to methods for vaccination of mammalian species
 CC with variants of E. coli FimH protein derived from different strains of
 CC E. coli. The vaccine composition or the antibody is useful for protecting
 CC against and treating an enterobacillus-related disease in a patient
 CC afflicted or at a risk of contracting the disease. In particular, the
 CC disease is a urinary tract or bladder infection. The disease is caused
 CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
 CC The present sequence is Escherichia coli strain EC56 FimH protein.
 CC
 XX
 SO Sequence 279 AA:
 Query Match 98.3%; Score 834; DB 23; Length 279;
 Best Local Similarity 98.1%; Pred. No. 6.9e-82;
 Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PVTNVCNGLVVDSTQIFCHNDYPERITTDVYTIQKRSAGVGNFSGTYKYGSSYPP 60
 DB 26 PVTNVCNGLVVDSTQIFCHNDYPERITTDVYTIQKRSAGVGNFSGTYKYGSSYPP 85
 QY 61 TTSETPRVYVNSRTDKPMPALYLTTPVSSAGLVIRAGSLIAYLLRQTNVNSDDFOFV 120
 DB 86 TTSETPRVYVNSRTDKPMPALYLTTPVSSAGLVIRAGSLIAYLLRQTNVNSDDFOFV 145
 QY 121 WNTIYANDVVPYTGCDVSARDVTYTLDPYRGSVPIPLTY 161
 DB 146 WNTIYANDVVPYTGCDVSARDVTYTLDPYRGSVPIPLTY 186
 RESULT 8
 AAE18433 standard; Protein: 279 AA.
 XX
 AC AAE18433:
 XX
 DT 07-MAY-2002 (first entry)
 DE
 DE Escherichia coli strain J96 FimH protein.
 XX
 KM FimH: immune response; antibacterial; enterobacillus-related disease;
 KM therapy; vaccine; urinary tract infection; bladder.
 XX
 OS Escherichia coli J96.
 XX
 PN WO200204496-A2.
 PD 17-JAN-2002.
 PF 06-JUL-2001; 2001MO-US21525.
 PR 07-JUL-2000; 2000US-216750P.
 PA (MEDT-) MEDIMMUNE INC.
 PI Langermann S, Revel A, Auguste C, Burlain J;
 DR WPI: 2002-171702/22.
 DR N-PSDB; AAD29372.
 PT New immunogenic polypeptide, useful as vaccine for protecting against
 PT an enterobacillus-related disease in a patient at risk of contracting
 PT such disease, e.g. urinary tract infection or a bladder infection -
 XX
 PS Claim 3; Page 90-91; 101pp; English.
 CC The invention relates to bacterial immunogenic agents for administration
 CC to humans and non-human animals to stimulate an immune response. The
 CC invention also relates to methods for vaccination of mammalian species

CC with variants of E. coli FimH protein derived from different strains of
 CC E. coli. The vaccine composition or the antibody is useful for protecting
 CC against and treating an enterobacillus-related disease in a patient
 CC afflicted or at a risk of contracting the disease. In particular, the
 CC disease is a urinary tract or bladder infection. The disease is caused
 CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
 CC The present sequence is Escherichia coli strain J96 FimH protein.
 CC
 XX
 SO Sequence 279 AA:
 Query Match 98.3%; Score 834; DB 23; Length 279;
 Best Local Similarity 98.1%; Pred. No. 6.9e-82;
 Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PVTNVCNGLVVDSTQIFCHNDYPERITTDVYTIQKRSAGVGNFSGTYKYGSSYPP 60
 DB 26 PVTNVCNGLVVDSTQIFCHNDYPERITTDVYTIQKRSAGVGNFSGTYKYGSSYPP 85
 QY 61 TTSETPRVYVNSRTDKPMPALYLTTPVSSAGLVIRAGSLIAYLLRQTNVNSDDFOFV 120
 DB 86 TTSETPRVYVNSRTDKPMPALYLTTPVSSAGLVIRAGSLIAYLLRQTNVNSDDFOFV 145
 QY 121 WNTIYANDVVPYTGCDVSARDVTYTLDPYRGSVPIPLTY 161
 DB 146 WNTIYANDVVPYTGCDVSARDVTYTLDPYRGSVPIPLTY 186
 RESULT 9
 AAR6769 standard; Protein: 300 AA.
 XX
 AC AAR6769:
 XX
 DT 15-MAR-1996 (first entry)
 DE
 DE FimH protein derived from E. coli clinical isolate CI#3.
 XX
 KM FimH: type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
 KM FimA; FimG; receptor binding site.
 XX
 OS Escherichia coli clinical isolate CI#3.
 XX
 PN WO9520657-A1.
 PD 03-AUG-1995.
 PF 27-JAN-1995; 95MO-DK00042.
 PR 27-JAN-1994; 94US-0187166.
 PA (GXBI-) GX BIOSYSTEMS AS.
 PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
 DR WPI: 1995-275442/36.
 DR N-PSDB; AAQ93071.
 PT Receptor specific bacterial adhesins - useful for targetting active
 PT compounds and microbial cells to locations of receptors
 XX
 PS Example 1; Page 44-45; 152pp; English.
 CC The sequences given in AAR6769-76 are FimH proteins from various E.
 CC coli clinical isolates. FimH is located at the tip of the type 1
 CC fimbriae and also intercalated at intervals in the fimbrial organelle.
 CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
 CC structures containing terminally located alpha-D-mannoside residues.

CC FimH contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison FimA and
CC the minor components FimF and FimG only have 2 cysteine residues.
CC The localisation of the cysteine residues in FimH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FimH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FimH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of a variant FimH adhesin which
CC may be useful for targeting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.

XX Sequence 300 AA;

Query Match 98.3%; Score 834; DB 16; Length 300;
Best Local Similarity 98.1%; Pred. No. 7.6e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PVTNNGQMLVVDLSTQIFCHNDYPETITDYVTLQSGSAYGVLNFSGTVKSGSSYFPF 60
DB 47 PAVNNGQMLVVDLSTQIFCHNDYPETITDYVTLQSGSAYGVLNFSGTVKSGSSYFPF 106
QY 61 TTSETPRVVYNSRTDKPMPVALYLTLPVSSAGGLVTKAGSLIAVLILRTNNYNSDDPQFV 120
DB 107 TTSETPRVVYNSRTDKPMPVALYLTLPVSSAGGLVTKAGSLIAVLILRTNNYNSDDPQFV 166
QY 121 WNIYANDVVPVPTGCGDVASARDVITVLPDYGSPVPIPLTVY 161
DB 167 WNIYANDVVPVPTGCGDVASARDVITVLPDYGSPVPIPLTVY 207

RESULT 10
AAR6745

ID AAR6745 standard; protein; 300 AA.

XX AAR6745;

DT 13-MAR-1996 (first entry).

DE FimH protein derived from E. coli K12 strain PC31.

KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;

KM FimA; FimF; FimG; receptor binding site.

OS Escherichia coli K12 strain PC31.

XX Key Location/Qualifiers

FT Peptide 1..21

FT /note= "Signal peptide"

FT Protein 22..300

FT /note= "Mature FimH"

XX W09520657-A1.

XX 03-AUG-1995.

XX 27-JAN-1995; 95WO-DK00042.

XX 27-JAN-1994; 94US-0187166.

XX (GABI-) GX BIOSYSTEMS AS.

PI Hasty DL, Klemm P, Molin S, Palleisen L, Sokurenko EV;

XX WPI; 1995-275442/36.

PT Receptor specific bacterial adhesins - useful for targeting active
XX compounds and microbial cells to locations of receptors

XX Example 1; Page 88-89; 152pp; English.

XX This sequence represents the FimH protein from E. coli K12 strain
CC PC31. FimH is located at the tip of the type 1 fimbriae and also
CC intercalated at intervals in the fimbrial organelle. Most forms of
CC the FimH adhesin target, and bind to, oligosaccharide structures
CC containing terminally located alpha-D-mannoside residues. FimH
CC contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison FimA and
CC the minor components FimF and FimG only have 2 cysteine residues.
CC The localisation of the cysteine residues in FimH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FimH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FimH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. This
CC sequence and those given in AAR6763-76 may be used in the production of
CC a variant FimH adhesin which may be useful for targeting active
CC compounds and microbial cells to locations comprising selected receptors
CC to which the adhesins bind.

XX Sequence 300 AA;

Query Match 98.3%; Score 834; DB 16; Length 300;
Best Local Similarity 98.1%; Pred. No. 7.6e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PVTNNGQMLVVDLSTQIFCHNDYPETITDYVTLQSGSAYGVLNFSGTVKSGSSYFPF 60
DB 47 PAVNNGQMLVVDLSTQIFCHNDYPETITDYVTLQSGSAYGVLNFSGTVKSGSSYFPF 106
QY 61 TTSETPRVVYNSRTDKPMPVALYLTLPVSSAGGLVTKAGSLIAVLILRTNNYNSDDPQFV 120
DB 107 TTSETPRVVYNSRTDKPMPVALYLTLPVSSAGGLVTKAGSLIAVLILRTNNYNSDDPQFV 166
QY 121 WNIYANDVVPVPTGCGDVASARDVITVLPDYGSPVPIPLTVY 161
DB 167 WNIYANDVVPVPTGCGDVASARDVITVLPDYGSPVPIPLTVY 207

RESULT 11

ID AAB16009 standard; protein; 300 AA.

XX AAB16009;

DT 05-OCT-2000 (first entry)

DE E. coli proliferation associated protein sequence SPQ ID NO:367.

KW Escherichia coli; E. coli; proliferation; inhibition; screening;

KM antimicrobial; bacterial growth; antisense therapy; antibacterial.

OS Escherichia coli.

XX W0200044906-A2.

XX 03-AUG-2000.

XX 27-JAN-2000; 2000WO-US02200.

XX 27-JAN-1999; 99US-0117405.

XX (ELIT-) ELITRA PHARM INC.

PI Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI; 2000-514822/46.

XX N-PSDB; AAA66015.

PT Novel polynucleotides and polypeptides associated with microorganism

PT proliferation, used to identify inhibitors of bacterial growth and
PT proliferation, for use in antisense therapy -
XX
PS Claim 11: Page 274-275; 316pp; English.
XX
CC AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
CC sequences derived from *Escherichia coli* which inhibit *E. coli*
CC proliferation. AAA65890 to AAA66055 and AAA65886 to AAA66040 represent
CC nucleotide and protein sequences associated with *E. coli* proliferation.
CC AAA66056 and AAA66057 represent primers used for sequencing *E. coli*
CC proliferation inhibiting nucleotide inserts in an example from the
CC present invention. Methods from the present invention can be used to
CC identify a proliferation-regulated gene in a microorganism, by contacting
CC a microorganism with a proliferation-regulated gene activity inhibitory
CC nucleic acid identified in another organism, and determining if
CC inhibition occurs in the second microorganism. The nucleic acid sequences
CC identified as being required for bacterial growth and proliferation, can
CC be used for antisense therapy for killing bacteria.
XX
SQ Sequence 300 AA:
Query Match 98.3%; Score 834; DB 21; Length 300;
Best Local Similarity 98.1%; Pred. No. 7.6e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PYYVNGNLVVDLSTQIFCHNDYPETITDVTYTLQSGSAGVLSNFGTYKSGSSYPFP 60
Db 47 PYYVNGNLVVDLSTQIFCHNDYPETITDVTYTLQSGSAGVLSNFGTYKSGSSYPFP 106
QY 61 TTSETPRVYVNSRTDKPMPVALYLTTPVSSAGLVAKGSLIAVLILRQTNVNSDDFOFV 120
Db 107 TTSETPRVYVNSRTDKPMPVALYLTTPVSSAGLVAKGSLIAVLILRQTNVNSDDFOFV 166
QY 121 WNIYANDVYVPTGGCDVSAKDVTYTLPDYRGSVPIPLTVY 161
Db 167 WNIYANDVYVPTGGCDVSAKDVTYTLPDYRGSVPIPLTVY 207
Db
RESULT 12
AAV59456
ID AAV59456 standard; peptide: 300 AA.
AC AAV59456;
XX
XX 29-MAR-2000 (first entry)
XX
XX E. coli PC31 FliH protein.
XX
XX Multifunctional adhesin protein; organic receptor; bioremediation;
XX biosorption; organic pollutant; herbicide; pesticide; toxic compound;
XX recycling; metal isolation; metal binding domain.
XX
XX *Escherichia coli*.
XX
XX WO9957276-A1.
XX
XX 11-NOV-1999.
XX
XX 21-APR-1999; 99WO-DK00223.
XX
XX 30-APR-1998; 98DK-0000598.
XX 01-MAY-1998; 98US-0083794.
XX
XX (GYRE-) GYRE LTD.
XX
XX Schembri MA, Klemm P;
XX
XX WPI, 2000-072233/06.
XX
XX Novel recombinant cells useful for bioremediation and recycling
XX processes -
XX
XX Claim 5; Page 5; 60pp; English.

XX
CC This sequence is the *E. coli* FliH protein, which is an adhesin protein.
CC The invention relates to a recombinant cell expressing a multifunctional
CC adhesin (MA) protein on its surface. The MA protein has at least one
CC binding domain (BD1) capable of binding to an organic receptor, and at
CC least one binding domain (BD2) not naturally present in the adhesin, and
CC can bind to a compound to which the naturally occurring adhesin protein
CC does not substantially bind. Cells of the invention are used as
CC bioremediation or biosorption means to separate undesired compounds such
CC as organic pollutants including herbicides and pesticides, or toxic
CC compounds such as heavy metals from the environment, or for isolating
CC precursors compounds such as metals for recycling purposes.
XX
SQ Sequence 300 AA:
Query Match 98.3%; Score 834; DB 21; Length 300;
Best Local Similarity 98.1%; Pred. No. 7.6e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PYYVNGNLVVDLSTQIFCHNDYPETITDVTYTLQSGSAGVLSNFGTYKSGSSYPFP 60
Db 47 PYYVNGNLVVDLSTQIFCHNDYPETITDVTYTLQSGSAGVLSNFGTYKSGSSYPFP 106
QY 61 TTSETPRVYVNSRTDKPMPVALYLTTPVSSAGLVAKGSLIAVLILRQTNVNSDDFOFV 120
Db 107 TTSETPRVYVNSRTDKPMPVALYLTTPVSSAGLVAKGSLIAVLILRQTNVNSDDFOFV 166
QY 121 WNIYANDVYVPTGGCDVSAKDVTYTLPDYRGSVPIPLTVY 161
Db 167 WNIYANDVYVPTGGCDVSAKDVTYTLPDYRGSVPIPLTVY 207
Db
RESULT 13
AAU77488
ID AAU77488 standard; protein: 300 AA.
AC AAU77488;
XX
XX 05-JUN-2002 (first entry)
XX
XX
XX *Escherichia coli* type I pilin protein, FliH.
XX
XX
XX Immune response; primate; immunoglobulin; urogenital tract infection;
XX FliH-FliH complex; FliH; IgG; human; urinary tract infection; URI;
XX bladder infection; kidney infection; Enterobacteriaceae; Bacteraea;
XX pregnant woman; diabetic; immunocompromised; HIV; cancer;
XX human immunodeficiency virus infection; end stage renal disease;
XX type I pilin protein; adhesin; FliH.
XX
XX *Escherichia coli* isolate J96.
XX
XX WO200215928-A1.
XX
XX 28-FEB-2002.
XX
XX 28-NOV-2000; 2000WO-US32398.
XX
XX 18-AUG-2000; 2000US-226146P.
XX
XX (MEDT-) MEDIMUNE INC.
XX
XX Langemann S, Ballou WR;
XX
XX WPI: 2002-280859/32.
XX N-PSDB: ABK11187.
XX
XX Stimulating immune response in a primate for preventing, treating
XX bacterial induced diseases such as diseases of urinary tract, by
XX administering bacterial adhesive proteins, preferably FliH-FliH
XX polypeptide complex -
XX
XX Claim 8; Page 89-90; 92pp; English.

CC The present invention relates to a method of inducing an immune response
CC in a primate. The response involves immunoglobulin (Ig) molecules that
CC bind a bacterial adhesin protein, preferably an attachment domain of
CC a type 1 pilin polypeptide (e.g. FimH) associated with a bacterium
CC causing urogenital tract infections (e.g. *Escherichia coli*). The method
CC comprises administering a purified FimH polypeptide, a FimC-FimH (FimCH)
CC complex, or immunogenic fragments of these. The method is useful for
CC inducing IgG molecules in a primate, especially human, to reduce or
CC prevent the incidence of urogenital tract infections, particularly
CC urinary tract infection (UTI), bladder infection, or kidney infection,
CC caused by a bacterium of the family Enterobacteriaceae, preferably
CC *E. coli*. The method can be used in a human subject that has suffered
CC more than two urogenital infections within one year, has asymptomatic
CC bacteriuria, is a pregnant woman or a diabetic, is immunocompromised, has
CC a human immunodeficiency virus (HIV) infection, has cancer, is in
CC remission from cancer, or is at risk for end stage renal disease. The
CC method is useful for vaccinating a primate against urogenital tract
CC infections, for treating or ameliorating the symptoms of urogenital
CC tract infections, and also for slowing or preventing progression of
CC a urinary tract infection into end stage renal disease. The present
CC sequence represents *E. coli* FimH protein.

SQ Sequence 300 AA;

Query Match 98.3%; Score 834; DB 23; Length 300;
Best Local Similarity 98.1%; Pred. No. 7.6e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAVNNGQNLVVDLSTQIFCHNDYPETITDYVTLQGSAYGVLSNFGTVKYSGSYPFP 60
DB 47 PAVNNGQNLVVDLSTQIFCHNDYPETITDYVTLQGSAYGVLSNFGTVKYSGSYPFP 106
QY 61 TTSETPRVVYNSRTDKPMPVALYLTLPVSSAGGLVYKAGSLAVLLRQTNNTNSDDFOFV 120
DB 107 TTSETPRVVYNSRTDKPMPVALYLTLPVSSAGGLVYKAGSLAVLLRQTNNTNSDDFOFV 166
QY 121 WNIYANDVVPVTGGCDVSARDVYVTLDPYRGSPVPIPLTV 161
DB 167 WNIYANDVVPVTGGCDVSARDVYVTLDPYRGSPVPIPLTV 207

RESULT 14

AAE18427
ID AAE18427 standard; Protein; 279 AA.

XX AAE18427;

DT 07-MAY-2002 (first entry)

XX Escherichia coli strain EC61 FimH protein.

XX FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.

OS *Escherichia coli* EC61.

XX Key Location/Qualifiers

FT Misc-difference 3 /note= "Encoded by TAT"

FT Misc-difference 176 /note= "Encoded by CCT"

FT Misc-difference 201 /note= "Encoded by ACC"

XX WO200204496-A2.

XX 17-JAN-2002.

XX 06-JUL-2001; 2001WO-US21525.

XX 07-JUL-2000; 2000US-216750P.

XX (MEDT-) MEDIMUNE INC.

XX Langermann S, Revel A, Auguste C, Burteln J;
XX WPT: 2002-171702/22.
DR N-PSDB: AAD29366.

XX New immunogenic polypeptide, useful as vaccine for protecting against
XX an enterobacillus-related disease in a patient at risk of contracting
XX such disease, e.g. urinary tract infection or a bladder infection
XX Claim 3; Fig 2; 101pp; English.

XX The invention relates to bacterial immunogenic agents for administration
XX to humans and non-human animals to stimulate an immune response. The
XX invention also relates to methods for vaccination of mammalian species
XX with variants of *E. coli* FimH protein derived from different strains of
XX *E. coli*. The vaccine composition or the antibody is useful for protecting
XX against and treating an enterobacillus-related disease in a patient
XX afflicted or at a risk of contracting the disease. In particular, the
XX disease is a urinary tract or bladder infection. The disease is caused
XX by a bacterium of the family Enterobacteriaceae, particularly *E. coli*.
XX The present sequence is *Escherichia coli* strain EC61 FimH protein.

SQ Sequence 279 AA;

Query Match 98.2%; Score 833; DB 23; Length 279;
Best Local Similarity 97.5%; Pred. No. 8.8e-82;
Matches 157; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAVNNGQNLVVDLSTQIFCHNDYPETITDYVTLQGSAYGVLSNFGTVKYSGSYPFP 60
DB 26 PAVNNGQNLVVDLSTQIFCHNDYPETITDYVTLQGSAYGVLSNFGTVKYSGSYPFP 85
QY 61 TTSETPRVVYNSRTDKPMPVALYLTLPVSSAGGLVYKAGSLAVLLRQTNNTNSDDFOFV 120
DB 86 TTSETPRVVYNSRTDKPMPVALYLTLPVSSAGGLVYKAGSLAVLLRQTNNTNSDDFOFV 145
QY 121 WNIYANDVVPVTGGCDVSARDVYVTLDPYRGSPVPIPLTV 161
DB 146 WNIYANDVVPVTGGCDVSARDVYVTLDPYRGSPVPIPLTV 186

RESULT 15

AAE18417
ID AAE18417 standard; Protein; 279 AA.

XX AAE18417;

DT 07-MAY-2002 (first entry)

XX Escherichia coli strain B228 FimH protein.

XX FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.

OS *Escherichia coli* B228.

XX Key Location/Qualifiers

FT Misc-difference 176 /note= "Encoded by CCT"

FT Misc-difference 203 /note= "Encoded by ACC"

XX WO200204496-A2.

XX 17-JAN-2002.

XX 06-JUL-2001; 2001WO-US21525.

XX 07-JUL-2000; 2000US-216750P.

XX (MEDT-) MEDIMUNE INC.

PI	Langermann S, Revel A, Auguste C, Burrell J;
XX	WP1: 2002-171702/22.
DR	N-PSDB: AAD29356.
XX	
PW	New immunogenic polypeptide, useful as vaccine for protecting against
PT	an enterobacillus-related disease in a patient at risk of contracting
PT	such disease, e.g. urinary tract infection or a bladder infection
XX	
PS	Claim 3; Fig 2; 101pp; English.
XX	
CC	The invention relates to bacterial immunogenic agents for administration
CC	to humans and non-human animals to stimulate an immune response. The
CC	invention also relates to methods for vaccination of mammalian species
CC	with variants of E. coli FimH protein derived from different strains of
CC	E. coli. The vaccine composition or the antibody is useful for protecting
CC	against and creating an enterobacillus-related disease in a patient
CC	afflicted or at a risk of contracting the disease. In particular, the
CC	disease is a urinary tract or bladder infection. The disease is caused
CC	by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC	The present sequence is Escherichia coli strain B328 FimH protein.
XX	
SO	Sequence 279 AA;
Query Match	96.1%; Score 832; DB 23; Length 279;
Best Local Similarity	97.5%; Pred. No. 1.1e-81;
Matches 157; Conservative 1; Mismatches 3; Indels 0; Gaps 0.	
OY	1 PYYNNGNLVNDLSQIFCHNDYPERITDYYLQSGAYGVLNPSGRTVYSGSSYPP 60
DB	26 PAVNNGNVLVDLSQIFCHNDYPERITDYYLQSGAYGVLNPSGRTVYSGSSYPP 85
OY	61 TTSETPRVYVNSRTDKPWEVALYLRPVSSAGSLVTKAGSLAVLILQRTNNYNSDDFOY 120
DB	86 TTSETPRVYVNSRTDKPWEVALYLRPVSSAGSLVTKAGSLAVLILQRTNNYNSDDFOY 145
OY	121 WNIYANNVDVYPPTGGCDVSAHDVYTLRPYRGSVPPLRYV 161
DB	146 WNIYANNVDVYPPTGGCDVSAHDVYTLRPYRGSVPPLRYV 186
RESULT 16	
AAEL18413	
ID	AAEL18413 standard; Protein; 279 AA.
XX	
AC	AAEL18413;
XX	
DT	07-MAY-2002 (first entry)
XX	
DE	Escherichia coli strain B210 FimH protein.
XX	
KW	FimH; immune response; antibacterial; enterobacillus-related disease;
XX	therapy; vaccine; urinary tract infection; bladder.
OS	Escherichia coli B210.
XX	
Key	Location/Qualifiers
FT	Misc-difference 141 /note= "Encoded by GGT"
FT	Misc-difference 176 /note= "Encoded by GGT"
FT	Misc-difference 201 /note= "Encoded by CCA"
FT	Misc-difference 203 /note= "Encoded by ACC"
FT	Misc-difference 203 /note= "Encoded by GGT"
XX	
PN	WO200204496-A2.
XX	
PD	17-JAN-2002.
XX	
PF	06-JUL-2001: 2001WO-US21525.
XX	
PR	07-JUL-2000: 2000US-216750P.

PA	(MEDI-) MEDIMUNE INC.
XX	
PI	Langermann S, Revel A, Auguste C, Bulein J;
XX	
DR	WPI; 2002-171702/22.
DR	N-PSDB; AAD29352.
XX	
PT	New immunogenic polypeptide, useful as vaccine for protecting against
PT	an enterobacillus-related disease in a patient at risk of contracting
PT	such disease, e.g. urinary tract infection or a bladder infection -
XX	
PS	Claim 3; Fig 2; 101pp; English.
XX	
CC	The invention relates to bacterial immunogenic agents for administration
CC	to humans and non-human animals to stimulate an immune response. The
CC	invention also relates to methods for vaccination of mammalian species of
CC	with variants of E. coli FimH protein derived from different strains of
CC	E. coli. The vaccine composition or the antibody is useful for protecting
CC	against and treating an enterobacillus-related disease in a patient
CC	afflicted or at a risk of contracting the disease. In particular, the
CC	disease is a urinary tract or bladder infection. The disease is caused
CC	by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC	The present sequence is Escherichia coli strain B210 FimH protein.
XX	
SQ	Sequence 279 AA:
Query Match	98.0%; Score 831; DB 23; Length 279;
Best Local Similarity	97.5%; Pred. No. 1.5e-81;
Matches 157; Conservative 1; Mismatches 3; Indels 0; Gaps 0;	
QY	1 PYYVNGQNLYVDLSQIFCHNDYPETITDYLQGSAYGVLSNFSGIVKXSGSSPP 60
DB	26 PYYVNGQNLYVDLSQIFCHNDYPETITDYLQGSAYGVLSNFSGIVKXSGSSPP 85
QY	61 TTSERPVRVYNSTRDKPMPVALYLPVVSAGGLVIKAGSLIAYLILRQTNNNSDDFOFV 120
DB	86 TTSERPVRVYNSTRDKPMPVALYLPVVSAGGLVIKAGSLIAYLILRQTNNNSDDFOFV 145
QY	121 WNIYANNDDVVYPTGGCDVASRDVTYTLPRYRGSVPIPLTV 161
DB	146 WNIYANNDDVVYPTGGCDVASRDVTYTLPRYRGSVPIPLTV 186
RESULT 17	
AAE1B432	
ID	AAE1B432 standard; Protein; 279 AA.
XX	
AC	AAE1B432;
XX	
DT	07-MAY-2002 (first entry)
XX	
DE	Escherichia coli strain G189 FimH protein.
XX	
KW	FimH; immune response; antibacterial; enterobacillus-related disease;
KW	therapy; vaccine; urinary tract infection; bladder.
XX	
OS	Escherichia coli G189.
XX	
FT	Key Location/Qualifiers
FT	Misc-difference 176 /note= "Encoded by CCR"
FT	Misc-difference 201 /note= "Encoded by ACC"
XX	
PN	WO200204496-A2.
PD	17-JAN-2002.
XX	
PF	06-JUL-2001; 2001WO-US21525.
XX	
PR	07-JUL-2000; 2000US-216750P.
XX	

PA (MEDI-) MEDIMUNE INC.
 XX Langermann S, Revel A, Auguste C, Burlein J;
 PI WPI: 2002-171702/22.
 DR N-PSDB; AAD29371.
 XX
 PT New immunogenic polypeptide, useful as vaccine for protecting against
 PR an enterobacillus-related disease in a patient at risk of contracting
 XX such disease, e.g. urinary tract infection or a bladder infection
 PS Claim 3; Fig 2; 101pp; English.
 XX
 CC The invention relates to bacterial immunogenic agents for administration
 CC to humans and non-human animals to stimulate an immune response. The
 CC invention also relates to methods for vaccination of mammalian species
 CC with variants of E. coli FimH protein derived from different strains of
 CC E. coli. The vaccine composition or the antibody is useful for protecting
 CC against and treating an enterobacillus-related disease in a patient
 CC afflicted or at a risk of contracting the disease. In particular, the
 CC disease is a urinary tract or bladder infection. The disease is caused
 CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
 CC The present sequence is Escherichia coli strain G189 FimH protein.
 XX
 SQ Sequence 279 AA;
 Query Match 98.0%; Score 831; DB 23; Length 279;
 Best Local Similarity 97.5%; Pred. No. 1.5e-81;
 Matches 157; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 PAVNVGQNLVVDLSTQIFCHNDYPETITDYTLQSGAYGVLNFSGTYKSGSSYFP 60
 Db 26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYTLQSGAYGVLNFSGTYKSGSSYFP 85
 QY 61 TTSETPRVVYNSRTPDKPMPALYLTLPVSSAGGLVYKAGSLAVLILRQTNMNSDDPQFV 120
 Db 86 TTSETPRVVYNSRTPDKPMPALYLTLPVSSAGGLVYKAGSLAVLILRQTNMNSDDPQFV 145
 QY 121 WNIYANDVVPVPGGCDVSARDVYTLDPDYGSPVPLTVY 161
 Db 146 WNIYANDVVPVPGGCDVSARDVYTLDPDYGSPVPLTVY 186

RESULT 18
 AAR76763 standard; protein: 300 AA.
 AC AAR76763;
 XX
 DT 15-MAR-1996 (first entry).
 XX
 DE FimH protein derived from E. coli clinical isolate KB21.
 XX
 KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
 KM FimA; FimF; FimG; receptor binding site.
 XX
 OS Escherichia coli clinical isolate KB21.
 XX
 FT Key Location/Qualifiers
 FT Peptide 1..21
 FT Protein /note="Signal peptide"
 FT Protein 22..300
 FT Protein /note="Mature FimH"
 XX
 PN WO9520657-A1.
 XX
 PD 03-AUG-1995.
 XX
 PF 27-JAN-1995; 95WO-DK00042.
 XX
 PR 27-JAN-1994; 94US-0187166.
 XX
 PA (GXBI-) GX BIOSYSTEMS AS.

XX
 PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
 XX WPI: 1995-275442/36.
 DR
 XX
 PT Receptor specific bacterial adhesins - useful for targetting active
 PR compounds and microbial cells to locations of receptors
 XX
 PS Example 1; Page 44-45; 152pp; English.
 XX
 CC The sequences given in AAR76763-76 are FimH proteins from various E.
 CC coli clinical isolates. FimH is located at the tip of the type 1
 CC fimbriae and also intercalated at intervals in the fimbrial organelle.
 CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
 CC structures containing terminally located alpha-D-mannoside residues.
 CC FimH contains 4 cysteine residues assumed to direct folding of the
 CC molecule into distinct functional domains. For comparison FimA and
 CC the minor components FimF and FimG only have 2 cysteine residues.
 CC The localisation of the cysteine residues in FimH points to a tandem
 CC arrangement of two ancestral genes. Similar amino acids can be
 CC found in similar positions in the two halves of the FimH protein. The
 CC "midway" point is located roughly around residue 150 in the mature
 CC protein. The two halves or domains of FimH have evolved differently
 CC with the N-terminal section becoming the domain harbouring the receptor
 CC binding site, whereas the C-terminal sector became the domain of the
 CC molecule required for integration into the fimbrial organelle. These
 CC sequences may be used in the production of a variant FimH adhesin which
 CC may be useful for targetting active compounds and microbial cells to
 CC locations comprising selected receptors to which the adhesins bind.
 XX
 SQ Sequence 300 AA;
 Query Match 97.9%; Score 830; DB 16; Length 300;
 Best Local Similarity 97.5%; Pred. No. 2.1e-81;
 Matches 157; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 PAVNVGQNLVVDLSTQIFCHNDYPETITDYTLQSGAYGVLNFSGTYKSGSSYFP 60
 Db 47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYTLQSGAYGVLNFSGTYKSGSSYFP 106
 QY 61 TTSETPRVVYNSRTPDKPMPALYLTLPVSSAGGLVYKAGSLAVLILRQTNMNSDDPQFV 120
 Db 107 TTSETPRVVYNSRTPDKPMPALYLTLPVSSAGGLVYKAGSLAVLILRQTNMNSDDPQFV 166
 QY 121 WNIYANDVVPVPGGCDVSARDVYTLDPDYGSPVPLTVY 161
 Db 167 WNIYANDVVPVPGGCDVSARDVYTLDPDYGSPVPLTVY 207

RESULT 19
 AAB47074 standard; protein: 279 AA.
 ID AAB47074
 AC AAB47074;
 XX
 DT 08-MAY-2001 (first entry)
 XX
 DE Adhesin protein, FimH.
 XX
 KW Donor; pilus protein; pilin; adhesin; vaccine; urinary epithelia;
 KM urinary tract infection; enterobacteriaceae.
 XX
 OS Escherichia coli.
 XX
 PN WO200104148-A2.
 XX
 PD 18-JAN-2001.
 XX
 PF 13-JUL-2000; 2000WO-US19066.
 XX
 PR 13-JUL-1999; 99US-0143582.
 XX
 PR 16-JUL-1999; 99US-0144359.
 XX
 PR 23-FEB-2000; 2000US-0184442.


```

XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Hultgren SJ, Pinkner JS, Sauer F, Barnhart M, Wakeman G, Knight S;
XX
DR WPI: 2001-138315/14.
XX
PT Immunogenic complexes and polypeptides for vaccinating against urinary
PT tract disease, comprises a pilus protein component and a bacterial
PT chaperrone -
XX
PS Disclosure: Fig 1; 92pp; English.
XX
CC This sequence may be used as the pilus protein in the immunogenic
CC complex of the invention. The complex comprises a pilus protein
CC component and a donor complement portion as part of the same amino
CC acid sequence or as non-covalently linked fragments of a complex
CC such that the correct conformation of the pilin is maintained. The
CC pilus protein component may be an adhesin or a pilin. Pilus associated
CC adhesins, such as FimH are relatively conserved proteins among
CC different species and strains of bacteria, therefore vaccines
CC incorporating the FimH antigen exhibit a broad spectrum of
CC protection compared with current pilus-fiber based vaccines. The
CC immunogenic complexes act by disrupting pilus-mediated attachment
CC of E. coli to urinary epithelia and may prevent or retard the
CC development of urinary tract infections. Vaccines containing the
CC complexes are useful for preventing urinary tract disease in a human
CC caused by the bacterium family enterobacteriaceae specifically
CC Escherichia coli and may also be used in treating the disease.
XX
SO Sequence 279 AA:

Query Match 97.8%; Score 829; DB 22; Length 279;
Best Local Similarity 97.5%; Pred. No. 2.4e-81;
Matches 157; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PNVNNGONLVVDLSTQIFCHNDYPERITDVTLORGSAVGGVLSNFGTKYSGSSYFP 60
DB 26 PNVNNGONLVVDLSTQIFCHNDYPERITDVTLORGSAVGGVLSNFGTKYSGSSYFP 85
QY 61 TTSETPRVYNSRTDKFMPALYLPVVSAGGLVYKAGSLIAVLIRQNNYNSDDFOY 120
DB 86 TTSETPRVYNSRTDKFMPALYLPVVSAGGLVYKAGSLIAVLIRQNNYNSDDFOY 145
QY 121 WNIYANDVYVPTGGCVSARDVTVTLPOYRGSPPIPLTVY 161
DB 146 WNIYANDVYVPTGGCVSARDVTVTLPOYRGSPPIPLTVY 186

RESULT 20
AAV72515
ID AAV72515 standard; Protein: 279 AA.
XX
AC AAV72515;
XX
DT 02-MAY-2001 (first entry)
XX
DE Escherichia coli FimH protein.
XX
FIMH: adhesin protein; type 1 pilus; mannose binding domain; MBD; COL;
KM collagen binding domain; prophylaxis; therapy; urinary tract infection;
KM URI; immunogen; passive immunotherapy; vaccine; antibacterial.
XX
OS Escherichia coli.
XX
FH
XX
FH Key Location/Qualifiers
FT Domain 1..156
FT Domain /label= lectin_binding_domain
FT Domain 1..20
FT Domain /label= MBD-1
FT Region /note= "Mannose binding domain-1"
FT /note= "beta strand 1"

```

```

FT Region
FT /note= "Beta strand 2"
FT 17..25
FT /note= "Beta strand 3"
FT 27
FT /label= FimC_chaperone_binding_site
FT 31
FT /label= FimC_chaperone_binding_site
FT 32..37
FT /note= "Beta strand 4a"
FT 38..41
FT /note= "3-10 helix"
FT 42..46
FT /note= "Beta strand 4b"
FT 46..54
FT /label= MBD-2
FT /note= "Mannose binding domain 2"
FT 50..80
FT /label= COL
FT /note= "Collagen binding domain: Forms a beta-sheet-
FT alpha-1-beta-sheet structure from strands 5, alpha-1
FT and 6"
FT 54..63
FT /note= "Beta strand 5"
FT 67..69
FT /note= "Alpha-1 helix"
FT 71..77
FT /note= "Beta strand 6"
FT 80..83
FT /note= "Beta strand 7"
FT 89..95
FT /note= "Beta strand 8"
FT 104..111
FT /note= "Beta strand 9"
FT 117
FT /label= FimC_chaperone_binding_site
FT 125..135
FT /note= "Beta strand 10"
FT 127..148
FT /label= MBD-3
FT /note= "Mannose binding domain 3"
FT 139..150
FT /note= "Beta strand 11"
FT 155
FT /label= FimC_chaperone_binding_site
FT 157..158
FT /label= FimC_chaperone_binding_site
FT 159..279
FT /label= Pilin_domain
FT 160
FT /label= FimC_chaperone_binding_site
FT 161..165
FT /note= "Beta strand A'"
FT 162..163
FT /label= FimC_chaperone_binding_site
FT 168..171
FT /note= "FimC_chaperone binding site; Serves as beta
FT strand A'"
FT 173
FT /label= FimC_chaperone_binding_site
FT 181
FT /label= FimC_chaperone_binding_site
FT 183
FT /label= FimC_chaperone_binding_site
FT 183..187
FT /note= "Beta strand B"
FT 191..192
FT /label= FimC_chaperone_binding_site
FT 192..197
FT /note= "beta strand C"
FT 196
FT /label= FimC_chaperone_binding_site
FT 198

```

FT	Binding-site	/label= FimC_chaperone_binding_site
FT	215	
FT	Region	/label= FimC_chaperone_binding_site
FT	221..226	
FT	Region	/note= "Beta strand D''"
FT	235..239	
FT	Region	/note= "Beta strand D''"
FT	251..258	
FT	Binding-site	/note= "beta strand E"
FT	266..279	
FT	Region	/label= FimC_chaperone_binding_site
FT	268..279	
FT	/note= "Beta strand F; contains conserved beta-zipper motif"	
XX		
PN	MO200105978-A1.	
XX		
PD	25-JAN-2001.	
XX		
PF	14-JUL-2000; 2000WO-US19402.	
XX		
PR	15-JUL-1999; 99US-0144016.	
PA	(MEDI-) MEDIMUNE INC.	
PI	Hultgren SJ, Langermann S;	
DR	WPI: 2001-159539/16.	
XX		
PT	Polypeptides useful as vaccines for prevention and/or treatment of diseases such as urinary tract infections, caused by Enterobacteriaceae, comprises mannose-binding domains derived from adhesin molecules -	
PS	Claim 1; Fig 3; 53pp; English.	
XX		
CC	The present sequence is FimH protein from Escherichia coli. FimH is an adhesin protein found in type 1 pili of bacteria of the family enterobacteriaceae, especially E. coli. The FimH protein comprises mannose-binding domains (MBDs) and collagen-binding domains (COL). The present invention relates to engineered polypeptides comprising one or more domains derived from FimH protein. These polypeptides are used to produce prophylactic vaccines which are useful for the prevention and/or treatment of diseases, such as urinary tract infection (UTI) caused by a bacterium of the family Enterobacteriaceae, especially E. coli in animals, in particular humans. They are useful as immunogens to stimulate the production of antibodies for use in passive immunotherapy, as a diagnostic reagent and as a reagent in other processes such as affinity chromatography. The antibodies of the novel polypeptides are also useful for research purposes for studying protein-leucin or collagen binding interactions.	
SQ	Sequence 279 AA:	
Query Match	97.8%; Score 829; DB 22; Length 279;	
Best Local Similarity	97.5%; Pred. No. 2,4e-81;	
Matches 157; Conservative	2; Mismatches 2; Indels 0; Gaps 0;	
OY	1 PVVNNGONLVLDLSTQIFCHNDYPETITDYVYLQGSAGAVLSNFSGTVKXSSGSPPP 60 PVTNNVGVNVLVDLSTQIFCHNDYPETITDYVYLQGSATGVLNSFSGTVKXSSSYP 85	
DB	26 PVTNNVGVNVLVDLSTQIFCHNDYPETITDYVYLQGSATGVLNSFSGTVKXSSSYP 85	
OY	61 TTSETPRRYYNRRTKPMFVALYLRPVSSAGGLVKRAGSLAVLLRQTNNYNSDFOFY 120 TTSETPRRYYNRRTKPMFVALYLRPVSSAGGAIVAKRAGSLAVLLRQTNNYNSDFOFY 145	
OY	121 WNIYANNDDVVPPTGGCDVSARDVYTLPIDRGSVPIPLTY 161 WNIYANNDDVVPPTGGCDVSARDVYTLPIDRGSVPIPLTY 186	
DB	146 WNIYANNDDVVPPTGGCDVSARDVYTLPIDRGSVPIPLTY 186	

ID	AAE18428	standard; protein; 279 AA.
XX	AAE18428;	
XX	07-MAY-2002	(first entry)
XX	Escherichia coli strain EC62	FimH protein.
XX	FimH; immune response; antibacterial; enterobacillus-related disease;	
XX	therapy; vaccine; urinary tract infection; bladder.	
XX	Escherichia coli EC62.	
XX	Key	Location/Qualifiers
XX	Misc-difference 176	/note= "Encoded by CCT"
XX	Misc-difference 201	/note= "Encoded by ACC"
XX	WO200204496-A2.	
XX	17-JAN-2002.	
XX	06-JUL-2001; 2001WO-US21525.	
XX	07-JUL-2000; 2000US-216750P.	
XX	(MEDI-) MEDIMUNE INC.	
XX	Langermann S, Revel A, Auguste C, Burlain J;	
XX	WPI: 2002-171702/22.	
XX	N-PSDB; AAD29367.	
XX	New immunogenic polypeptide, useful as vaccine for protecting against	
XX	an enterobacillus-related disease in a patient at risk of contracting	
XX	such disease, e.g. urinary tract infection or a bladder infection	
XX	Claim 3; Fig 2; 101pp; English.	
XX	The invention relates to bacterial immunogenic agents for administration	
XX	to humans and non-human animals to stimulate an immune response. The	
XX	invention also relates to methods for vaccination of mammalian species	
XX	with variants of E. coli FimH protein derived from different strains of	
XX	E. coli. The vaccine composition or the antibody is useful for protecting	
XX	against and treating an enterobacillus-related disease in a patient	
XX	afflicted or at a risk of contracting the disease. In particular, the	
XX	disease is a urinary tract or bladder infection. The disease is caused	
XX	by a bacterium of the family Enterobacteriaceae, particularly E. coli.	
XX	The present sequence is Escherichia coli strain EC62 FimH protein.	
XX	Sequence 279 AA;	
XX	Query Match 97.8%; Score 829; DB 23; Length 279;	
XX	Best Local Similarity 96.9%; Pred. No. 2.4e-81;	
XX	Matches 156; Conservative 3; Mismatches 2; Indels 0; Gaps 0;	
QY	1 PAVNVNGQNLVLDLSTQIFCHNDYPETITDVTLLQKRSAYGVLNFSGTVKYSGSSYPP	60
DB	26 PAVNVNGQNLVLDLSTQIFCHNDYPETITDVTLLQKRSAYGVLNFSGTVKYSGSSYPP	85
QY	61 TTSETPRVVYNNRTKPKPWPVALYILRPVSSAGGLVIRKAGLLAVLLIRQTNNNNSDDFOCV	120
DB	86 TTSETPRVVYNNRTKPKPWPVALYILRPVSSAGGLVIRKAGLLAVLLIRQTNNNNSDDFOCV	145
QY	121 WNIYANDVYVPTGGCDVSARDVYVTLTLPDYGSGVPIPLTIV	161
DB	146 WNIYANDVYVPTGGCDVSARDVYVTLTLPDYGSGVPIPLTIV	186

ID		AAEI8430 standard; Protein: 279 AA.
XX AC	AAE18430:	
XX DT	07-MAY-2002	(first entry)
XX DE	Escherichia coli strain EC89 FimH protein.	
XX KM	FIMH; immune response; antibacterial; enterobacillus-related disease;	
XX OS	therapy; vaccine; urinary tract infection; bladder.	
XX	Escherichia coli EC89.	
PX FH	Key	Location/Qualifiers
FT FT	Misc-difference 176	/note= "Encoded by CCR"
FT FT	Misc-difference 201	/note= "Encoded by ACC"
FT FT	Misc-difference 226	/note= "Encoded by GCG"
FT FT	Misc-difference 227	/note= "Encoded by CGC"
FT FT	Misc-difference 231	/note= "Encoded by GTT"
FT FT	Misc-difference 232	/note= "Encoded by ATT"
XX XX	WO200204496-A2.	
XX PD	17-JAN-2002.	
XX PF	06-JUL-2001; 2001MO-US21525.	
PR PR	07-JUL-2000; 2000US-216750P.	
PA PA	(MEDI-) MEDIMUNE INC.	
PX PI	Langermann S, Revel A, Auguste C, Burel J:	
DR DR	WPI: 2002-171702/22.	
XX N-PSDB:	AAD29369.	
PT PT	New immunogenic polypeptide, useful as vaccine for protecting against	
PT PT	an enterococcus-related disease in a patient at risk of contracting	
XX PT	such disease, e.g. urinary tract infection or a bladder infection -	
XX PS	Claim 3; Fig 2; 10pp; English.	
CC CC	The invention relates to bacterial immunogenic agents for administration	
CC CC	to humans and non-human animals to stimulate an immune response. The	
CC CC	invention also relates to methods for vaccination of mammalian species of	
CC CC	with variants of E. coli FimH protein derived from different strains of	
CC CC	E. coli. The vaccine composition or the antibody is useful for protecting	
CC CC	against and treating an enterococcus-related disease in a patient	
CC CC	afflicted or at a risk of contracting the disease. In particular, the	
CC CC	disease is a urinary tract or bladder infection. The disease is caused	
CC CC	by a bacterium of the family Enterobiaceae, particularly E. coli.	
XX SQ	The present sequence is Escherichia coli strain EC89 FimH protein.	
SQ Sequence	279 AA:	
Query Match	97.8%; Score 829; DB 23; Length 279;	
Best local Similarity	96.9%; Pred. No. 2, 4e-81;	
Matches 156; Conservative	3; Mismatches 2; Indels 0; Gaps	0
DG QY	1 PVTNVCNTLVDTSTOIFCHNDYPETITTDVYLQRGSATGVLSNFSGTYKYGSSPYPP 60 : 26 PAVNWGNLTVDDISTOFCHNDYPEITTDVYLQRSAGVGVLSPFSGYKYNGSSPYPP 85	
DG QY	61 TTSETPRVVYNSTRDKMPVPALTLTPVSSAGGLVKRAGSLIAVLINQTNNNSDPEFV 120 86 TTSERPVIYNSTRDKMPVPALTILTFPVSSAGVAIKRAGSLIANVLIRKTNNYSDDPEFV 145	

[illegible]

Best Local Similarity 97.5%; Pred. No. 2,6e-81;
Matches 157; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PVTNNGQNLVVDLSTQIFCHNDYPTETITDYVTLQSGSAYGVLNFSGTAKYSSGSSYFPF 60
DB 26 PVTNNGQNLVVDLSTQIFCHNDYPTETITDYVTLQSGSAYGVLNFSGTAKYSSGSSYFPF 85
QY 61 TTSEPRVYVNSRTDKPWPVALYLTLPVSSAGGLYIKAGSLIAVLILRQTNMYNSDDFOFV 120
DB 86 TTSEPRVYVNSRTDKPWPVALYLTLPVSSAGGLYIKAGSLIAVLILRQTNMYNSDDFOFV 145
QY 121 WNIYANDVYVPTGGCDVSARDVYVTLDPYRGSVPIPLTVY 161
DB 146 WNIYANDVYVPTGGCDVSARDVYVTLDPYRGSVPIPLTVY 186

RESULT 24

AAB47072
ID AAB47072 standard; Protein; 304 AA.

AC AAB47072;

DT 08-MAY-2001 (first entry)

DE Immunogenic complex: FimH-Linker-G1 beta-strand of FimC.

KW Donor: pilus protein; pilin; adhesin; vaccine; urinary epithelia;

KM urinary tract infection; enterobacteriaceae.

OS Chimeric - Escherichia coli.

OS Chimeric - Synthetic.

PH Key Location/Qualifiers

FT 1..279

FT /label= "FimH"

FT /note= "Pilus protein component"

FT Peptide

FT 280..289

FT /note= "Linker"

FT 290..304

FT /label= "G1 beta-strand of FimC"

FT /note= "Donor strand component"

PN WO200104148-A2.

PD 18-JAN-2001.

PF 13-JUL-2000; 2000WO-US19066.

XX 13-JUL-1999; 990S-0143582.

PR 16-JUL-1999; 990S-0144359.

PR 23-FEB-2000; 2000US-0184442.

XX (MEDI-) MEDIMUNE INC.

PA Hultgren SJ, Pinkner JS, Sauer F, Barnhart M, Waksman G, Knight S;

PI WPI; 2001-138315/14.

DR Immunogenic complexes and polypeptides for vaccinating against urinary

PT tract disease, comprises a pilus protein component and a bacterial

PT chaperone -

XX Claim 19; Page 80-81; 92pp; English.

PS This sequence represents the immunogenic complex of the invention.

XX The complex comprises a pilus protein component and a donor complement

CC portion as part of the same amino acid sequence attached through an

CC amino acid linker. The linker is composed of a sequence which

CC readily forms a loop such that the donor strand can loop back towards

CC the pilus protein and form an anti-parallel structure. Pilus associated

CC adhesins, such as FimH are relatively conserved proteins among

CC different species and strains of bacteria, therefore vaccines

CC incorporating the FimH antigen exhibit a broad spectrum of

CC protection compared with current pilus-fiber based vaccines. The
CC immunogenic complexes act by disrupting pilus-mediated attachment
CC of E. coli to urinary epithelia and may prevent or retard the
CC development of urinary tract infections. Vaccines containing the
CC complexes are useful for preventing urinary tract disease in a human
CC caused by the bacterium family enterobacteriaceae specifically
CC Escherichia coli and may also be used in treating the disease.

SO Sequence 304 AA;

Query Match 97.8%; Score 829; DB 22; Length 304;

Best Local Similarity 97.5%; Pred. No. 2,7e-81;

Matches 157; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PVTNNGQNLVVDLSTQIFCHNDYPTETITDYVTLQSGSAYGVLNFSGTAKYSSGSSYFPF 60
DB 26 PVTNNGQNLVVDLSTQIFCHNDYPTETITDYVTLQSGSAYGVLNFSGTAKYSSGSSYFPF 85
QY 61 TTSEPRVYVNSRTDKPWPVALYLTLPVSSAGGLYIKAGSLIAVLILRQTNMYNSDDFOFV 120
DB 86 TTSEPRVYVNSRTDKPWPVALYLTLPVSSAGGLYIKAGSLIAVLILRQTNMYNSDDFOFV 145
QY 121 WNIYANDVYVPTGGCDVSARDVYVTLDPYRGSVPIPLTVY 161
DB 146 WNIYANDVYVPTGGCDVSARDVYVTLDPYRGSVPIPLTVY 186

RESULT 25

AAR76768
ID AAR76768 standard; protein; 300 AA.

AC AAR76768;

DT 15-MAR-1996 (first entry)

DE FimH protein derived from E. coli clinical isolate CSH-50.

KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;

KM FimH; FimF; FimG; receptor binding site.

OS Escherichia coli clinical isolate CSH-50.

PH Key Location/Qualifiers

FT 1..21

FT /note= "Signal peptide"

FT Protein

FT 22..300

FT /note= "Mature FimH"

PN WO9520657-A1.

PD 03-AUG-1995.

PF 27-JAN-1995; 95WO-DK00042.

XX 27-JAN-1994; 94US-0187166.

PR (GXBL-) GX BIOSYSTEMS AS.

PA Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;

PI WPI; 1995-275442/36.

DR Receptor specific bacterial adhesins - useful for targeting active

PT compounds and microbial cells to locations of receptors

PT Example 1; Page 44-45; 152pp; English.

PS The sequences given in AAR76763-76 are FimH proteins from various E.

XX coli clinical isolates. FimH is located at the tip of the type 1

CC fimbriae and also intercalated at intervals in the fimbrial organelle.

CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide

CC structures containing terminally located alpha-D-mannoside residues.

CC FimH contains 4 cysteine residues assumed to direct folding of the

CC molecule into distinct functional domains. For comparison FlmH and
CC the minor components FlmF and FlmG only have 2 cysteine residues.
CC The localization of the cysteine residues in FlmH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FlmH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FlmH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of a variant FlmH adhesin which
CC may be useful for targeting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.

XX Sequence 300 AA:

SQ Query Match 97.6%; Score 828; DB 16; Length 300;

Best Local Similarity 97.5%; Pred. No. 3.4e-81;
Matches 157; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 PYYVNGNLVVDLSTQIFCHNDYPTITDYVTLQKRSAYGVLNFGSTVKYSGSSYPP 60
DB 47 PYYVNGNLVVDLSTQIFCHNDYPTITDYVTLQKRSAYGVLNFGSTVKYSGSSYPP 106
QY 61 TTSETPRVYNSRTDKPWPALYLTPEVSSAGVIAKGSILAVLILRQTNNNNSDDFOFY 120
DB 107 TTSETPRVYNSRTDKPWPALYLTPEVSSAGVIAKGSILAVLILRQTNNNNSDDFOFY 166
QY 121 WNIYANDVYVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
DB 167 WNIYANDVYVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 207

RESULT 26

AAE18429 ID AAE18429 standard; Protein: 279 AA.

XX AAE18429;

AC 07-MAY-2002 (first entry)

XX Escherichia coli strain EC80 FlmH protein.

XX FlmH: immune response; antibacterial; enterobacillus-related disease;

KW therapy; vaccine; urinary tract infection; bladder.

XX Escherichia coli EC80.

OS Escherichia coli EC80.

XX Key Location/Qualifiers
FH MISC-difference 27
FT MISC-difference 165 /note- "Encoded by GTC"
FT MISC-difference 171 /note- "Encoded by GCT"
FT MISC-difference 171 /note- "Encoded by ACT"
FT MISC-difference 269 /note- "Encoded by CAA"
FT /note- "Encoded by CAA"

XX WO200204496-A2.

XX 17-JAN-2002.

XX 06-JUL-2001; 2001MO-US21525.

XX 07-JUL-2000; 2000US-216750P.

XX (MEDI-) MEDIMUNE INC.

XX Langermann S, Revel A, Auguste C, Burteln J;

XX WPI: 2002-171702/22.

XX N-PSDB: AAD29368.

XX

PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection

PS Claim 3; Fig 2; 101pp; English.

XX The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FlmH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC80 FlmH protein.

XX Sequence 279 AA:

SQ Query Match 97.5%; Score 827; DB 23; Length 279;

Best Local Similarity 96.9%; Pred. No. 3.9e-81;
Matches 156; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 PYYVNGNLVVDLSTQIFCHNDYPTITDYVTLQKRSAYGVLNFGSTVKYSGSSYPP 60
DB 26 PAVVNGNLVVDLSTQIFCHNDYPTITDYVTLQKRSAYGVLNFGSTVKYSGSSYPP 85
QY 61 TTSETPRVYNSRTDKPWPALYLTPEVSSAGVIAKGSILAVLILRQTNNNNSDDFOFY 120
DB 86 TTSETPRVYNSRTDKPWPALYLTPEVSSAGVIAKGSILAVLILRQTNNNNSDDFOFY 145
QY 121 WNIYANDVYVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
DB 146 WNIYANDVYVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186

RESULT 27

AAE18420 ID AAE18420 standard; Protein: 280 AA.

XX AAE18420;

AC 07-MAY-2002 (first entry)

XX Escherichia coli strain B242 FlmH protein.

XX FlmH: immune response; antibacterial; enterobacillus-related disease;

KW therapy; vaccine; urinary tract infection; bladder.

XX Escherichia coli B242.

XX Key Location/Qualifiers
FH MISC-difference 176
FT MISC-difference 201 /note- "Encoded by CCT"
FT MISC-difference 201 /note- "Encoded by ACC"
FT MISC-difference 279..280 /note- "Encoded by CAA"
FT /note- "Encoded by CAA"

XX WO200204496-A2.

XX 17-JAN-2002.

XX 06-JUL-2001; 2001MO-US21525.

XX 07-JUL-2000; 2000US-216750P.

XX (MEDI-) MEDIMUNE INC.

XX Langermann S, Revel A, Auguste C, Burteln J;

XX WPI: 2002-171702/22.

XX N-PSDB: AAD29359.

XX

```
XX New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
XX
XX Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain B242 FimH protein.
CC
SQ Sequence 280 AA;
Query Match 97.5%; Score 827; DB 23; Length 280;
Best Local Similarity 96.9%; Pred. No. 4e-81;
Matches 156; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 PVTNNGQNLVVDLSTQIFCHNDYPTETIDYVTLQGSAYGVLNFSGTVKYSGSSYFP 60
DB 26 PAVNNGQNLVVDLSTQIFCHNDYPTETIDYVTLQGSAYGVLNFSGTVKYSGSSYFP 85
QY 61 TTSETPRVYNSRTDKPMPVALYTLTPVSSAGGLVKAGSLAVLLRQTNNNYNSDDPQFV 120
DB 86 TTSETPRVYNSRTDKPMPVALYTLTPVSSAGGLVKAGSLAVLLRQTNNNYNSDDPQFV 145
QY 121 WNIYANNDDVVPPTGCDVSARDVTVTLDPYRGSVPIPLTVY 161
DB 146 WNIYANNDDVVPPTGCDVSARDVTVTLDPYRGSVPIPLTVY 186
RESULT 28
AAE18415
ID AAE18415 standard; Protein; 279 AA.
XX
AC AAE18415;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain B217 FimH protein.
XX
KM FimH; immune response; antibacterial; enterobacillus-related disease;
KM therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli B217.
XX
FH Key Location/Qualifiers
FT Misc-difference 62 /note= "Encoded by TCG"
FT Misc-difference 70 /note= "Encoded by AAT"
FT Misc-difference 78 /note= "Encoded by AGT"
FT Misc-difference 176 /note= "Encoded by CCG"
FT Misc-difference 201 /note= "Encoded by ACC"
FT Misc-difference 234 /note= "Encoded by ACG"
FT Misc-difference 244 /note= "Encoded by CGG"
FT /note= "Encoded by CGG"
XX
XX WO200204496-A2.
XX
XX 17-JAN-2002.
XX
XX 06-JUL-2001; 2001WO-US21525.
XX
```

```
XX
PR 07-JUL-2000; 2000US-216750P.
XX
XX (MEDI-) MEDIMUNE INC.
XX
XX Langermann S, Revel A, Auguste C, Burlein J;
XX
XX WPI: 2002-171702/22.
XX
XX N-PSTDB; AAD29354.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
XX
XX Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain B217 FimH protein.
CC
SQ Sequence 279 AA;
Query Match 97.4%; Score 826; DB 23; Length 279;
Best Local Similarity 96.3%; Pred. No. 5.1e-81;
Matches 155; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 1 PVTNNGQNLVVDLSTQIFCHNDYPTETIDYVTLQGSAYGVLNFSGTVKYSGSSYFP 60
DB 26 PAVNNGQNLVVDLSTQIFCHNDYPTETIDYVTLQGSAYGVLNFSGTVKYSGSSYFP 85
QY 61 TTSETPRVYNSRTDKPMPVALYTLTPVSSAGGLVKAGSLAVLLRQTNNNYNSDDPQFV 120
DB 86 TTSETPRVYNSRTDKPMPVALYTLTPVSSAGGLVKAGSLAVLLRQTNNNYNSDDPQFV 145
QY 121 WNIYANNDDVVPPTGCDVSARDVTVTLDPYRGSVPIPLTVY 161
DB 146 WNIYANNDDVVPPTGCDVSARDVTVTLDPYRGSVPIPLTVY 186
RESULT 29
AAE18416
ID AAE18416 standard; Protein; 279 AA.
XX
AC AAE18416;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain B223 FimH protein.
XX
KM FimH; immune response; antibacterial; enterobacillus-related disease;
KM therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli B223.
XX
FH Key Location/Qualifiers
FT Misc-difference 176 /note= "Encoded by CCG"
FT Misc-difference 201 /note= "Encoded by ACC"
FT /note= "Encoded by ACC"
XX
XX WO200204496-A2.
XX
XX 17-JAN-2002.
XX
XX 06-JUL-2001; 2001WO-US21525.
XX
```

XX	07-JUL-2000; 2000US-216750P.
PA	(MEDI-) MEDIMMUNE INC.
PI	Langermann S, Revel A, Auguste C, Burieln J;
XX	
DR	WPI; 2002-171702/22.
XX	N-PSDB; AAD29355.
PT	New immunogenic polypeptide, useful as vaccine for protecting against
PT	an enterobacillus-related disease in a patient at risk of contracting
PT	such disease, e.g. urinary tract infection or a bladder infection -
XX	
PS	Claim 3; Fig 2; 101pp; English.
XX	
CC	The invention relates to bacterial immunogenic agents for administration
CC	to humans and non-human animals to stimulate an immune response. The
CC	invention also relates to methods for vaccination of mammalian species
CC	with variants of E. coli FimH protein derived from different strains of
CC	E. coli. The vaccine composition or the antibody is useful for protecting
CC	against and treating an enterobacillus-related disease in a patient
CC	afflicted or at a risk of contracting the disease. In particular, the
CC	disease is a urinary tract or bladder infection. The disease is caused
CC	by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC	The present sequence is Escherichia coli strain B23 FimH protein.
XX	
SQ	Sequence 279 AA;
	Query Match 97.4%; Score 826; DB 23; Length 279;
	Best Local Similarity 96.3%; Pred. No. 5,1e-81;
	Matches 155; Conservative 4; Mismatches 2; Indels 0; Gaps 0.
OY	1 PAVNNGONLVDLSIQIFCHNDYPETITDVTYLORGSAVAGVLSNFSGTVKXGSSRPP 60
Db	26 PAVNNGONLVDSLISQIFCHNDYPETITDVTYLORGSAVAGVLSNFGTVKXGSSRPP 85
OY	61 TTSEPRVVYNSRTDKPMFVALYLTPVSSAGGLVIKAGSLIAYLIKRQNNNSDDFOY 120
Db	86 TTSEPRVVYNSRTDKPMFVALYLTPVSSAGGVAIKAGSLIAYLIKRQNNNSDDFOY 145
OY	121 NNIYANNPDVVPPTGGCDVSARDVTYTLPRYGSVPRIPLTY 161
Db	146 NNIYANNPDVVPPTGGCDVSARDVTYTLPRYGSVPRIPLTY 186
	RESULT 30
ID	AAE18434
AC	AAE18434 standard; Protein; 279 AA.
XX	
DT	07-MAY-2002 (first entry)
DE	
XX	
FamH:	Escherichia coli strain NU14 FimH protein.
KM	FamH: Immune response; antibacterial; enterobacillus-related disease;
OS	therapy; vaccine; urinary tract infection; bladder.
XX	
ES	Escherichia coli NU14.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 176
FT	Misc-difference 201 /note= "Encoded by CCI"
FT	Misc-difference 201 /note= "Encoded by ACC"
XX	
PN	WO200204496-A2.
PD	17-JAN-2002.
Pf	06-JUL-2001; 2001WO-US21525.
Pr	07-JUL-2000; 2000US-216750P.

[illegible]

PA	(MED1-) MEDIMMUNE INC.
XX	Langermann S, Revel A, Auguste C, Burlain J;
XX	WPI; 2002-171702/22.
DR	N-PSDB; AAD29360.
XX	
XX	New immunogenic polypeptide, useful as vaccine for protecting against
PT	an enterobacillus-related disease in a patient at risk of contracting
FT	such disease, e.g. urinary tract infection or a bladder infection -
XX	
PS	Claim 3; Fig 2; 101pp; English.
XX	
CC	The invention relates to bacterial immunogenic agents for administration
CC	to humans and non-human animals to stimulate an immune response. The
CC	invention also relates to methods for vaccination of mammalian species
CC	with variants of E. coli FimH protein derived from different strains of
CC	E. coli. The vaccine composition or the antibody is useful for protecting
CC	against and treating an enterobacillus-related disease in a patient
CC	afflicted or at a risk of contracting the disease. In particular, the
CC	disease is a urinary tract or bladder infection. The disease is caused
CC	by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC	The present sequence is Escherichia coli strain DS17 FimH protein.
XX	
SO	Sequence 279 AA:
Query Match	97.3%; Score 825; DB 23; Length 279;
Best Local Similarity	96.3%; Pred. No. 6..5e-81;
Matches 155; Conservative 3; Mismatches 3; Indels 0; Gaps 0	
OY	1 PYYVNGQLVVDLSQIFCHNDYPETITDYVLQRGSATGVLNFGVKISGSSYPP 60
Db	26 PAVNVGQNLVPLSLQIFCHNDYPETITDYVLQRGSAGVLSFSGTKYNGSSYPP 85
OY	61 TTSSEPRVVYNSRTKRPMPVALYLTPVSSAGGLVKRAGSLIVLLIROTNNYNSDDEFQV 120
Db	86 TTSEPRVVYNSRTKRPMPVALYLTPVSSAGGVAKRAGSLIVLLIROTNNYNSDDPFQV 145
OY	121 WNIYANDVVPVTGGCDVSARDVTVTLDPYRGSVPIPLTV 161
Db	146 WNIYANDVVPVTGGCDASARPDVTVTLDPYRGSVPIPLTV 186
RESULT 32	
ID	AAR6771 standard; protein; 300 AA.
AC	AAR6771;
XX	
DT	15-MAR-1996 (first entry)
XX	
DE	FimH protein derived from E. coli clinical isolate KS-54.
XX	
KW	FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KM	FimA; FimF; FimG; receptor binding site.
OS	Escherichia coli clinical isolate KS-54.
XX	
FT	Key location/qualifiers
FT	Peptide 1..21
FT	/note= "Signal peptide"
FT	Protein 22..300
FT	/note= "Mature FimH"
PN	WO9520657-A1.
XX	
PD	03-AUG-1995.
XX	
PF	27-JAN-1995; 95MO-DK00042.
XX	
PR	27-JAN-1994; 94US-0187166.
XX	
PA	(GXBI-) GX BIOSYSTEMS AS.

[illegible]

PD 03-AUG-1995.
 XX
 PF 27-JAN-1995; 95WO-DK00042.
 XX
 PR 27-JAN-1994; 94US-0187166.
 XX
 PA (GXBI-) GX BIOSYSTEMS AS.
 XX
 PI Hasty DL, Klemm P, Molin S, Palleesen L, Sokurenko EV;
 XX
 DR WPI: 1995-275442/36.
 DR N-PSDB; AAQ93067.
 XX
 PT Receptor specific bacterial adhesins - useful for targeting active
 PT compounds and microbial cells to locations of receptors
 XX
 PS Example 1; Page 44-45; 152pp; English.
 XX
 CC The sequences given in AAR76763-76 are FimH proteins from various E.
 CC coli clinical isolates. FimH is located at the tip of the type 1
 CC fimbriae and also intercalated at intervals in the fimbrial organelle.
 CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
 CC structures containing terminally located alpha-D-mannoside residues.
 CC FimH contains 4 cysteine residues assumed to direct folding of the
 CC molecule into distinct functional domains. For comparison FimA and
 CC the minor components FimF and FimG only have 2 cysteine residues.
 CC The localisation of the cysteine residues in FimH points to a tandem
 CC arrangement of two ancestral genes. Similar amino acids can be
 CC found in similar positions in the two halves of the FimH protein. The
 CC "midway" point is located roughly around residue 150 in the mature
 CC protein. The two halves or domains of FimH have evolved differently
 CC with the N-terminal section becoming the domain harbouring the receptor
 CC binding site, whereas the C-terminal sector became the domain of the
 CC molecule required for integration into the fimbrial organelle. These
 CC sequences may be used in the production of a variant FimH adhesin which
 CC may be useful for targeting active compounds and microbial cells to
 CC locations comprising selected receptors to which the adhesins bind.
 CC
 XX
 SQ Sequence 300 AA;
 Query Match 97.3%; Score 825; DB 16; Length 300;
 Best Local Similarity 96.9%; Pred. No. 7.2e-81;
 Matches 156; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 PVAVNGNLVVDLSTQIFCHNDYPERITTDVYTLQRGSAVGGVSNFGTVKYGSSVPP 60
 DB 47 PAVAVNGNLVVDLSTQIFCHNDYPERITTDVYTLQRGSAVGGVSNFGTVKYGSSVPP 106
 QY 61 TTSETPRVYNSRTDKPMPALYLTVPVSSAGLVKAGSLIAVLILQTNNNYSDDFQFV 120
 DB 107 TTSETPRVYNSRTDKPMPALYLTVPVSSAGLVKAGSLIAVLILQTNNNYSDDFQFV 166
 QY 121 WNIYANDVYVPTGGCDVSAHDVTVTLDPYRGSVPIPLTVY 161
 DB 167 WNIYANDVYVPTGGCDVSAHDVTVTLDPYRGSVPIPLTVY 207
 RESULT 34
 AAR76774
 ID AAR76774 standard; protein; 300 AA.
 AC AAR76774;
 XX
 DT 15-MAR-1996 (first entry)
 XX
 DE FimH protein derived from E. coli clinical isolate MJ#31-3.
 XX
 KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
 KW FimA; FimG; receptor binding site.
 OS Escherichia coli clinical isolate MJ#31-3.
 XX
 FH Key Location/Qualifiers

FT Peptide 1..21
 FT /note= "Signal peptide"
 FT Protein 22..300
 FT /note= "Mature FimH"
 PN W09520657-A1.
 XX
 PD 03-AUG-1995.
 XX
 PF 27-JAN-1995; 95WO-DK00042.
 XX
 PR 27-JAN-1994; 94US-0187166.
 XX
 PA (GXBI-) GX BIOSYSTEMS AS.
 XX
 PI Hasty DL, Klemm P, Molin S, Palleesen L, Sokurenko EV;
 XX
 DR WPI: 1995-275442/36.
 DR N-PSDB; AAQ93068.
 XX
 PT Receptor specific bacterial adhesins - useful for targeting active
 PT compounds and microbial cells to locations of receptors
 XX
 PS Example 1; Page 44-45; 152pp; English.
 XX
 CC The sequences given in AAR76763-76 are FimH proteins from various E.
 CC coli clinical isolates. FimH is located at the tip of the type 1
 CC fimbriae and also intercalated at intervals in the fimbrial organelle.
 CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
 CC structures containing terminally located alpha-D-mannoside residues.
 CC FimH contains 4 cysteine residues assumed to direct folding of the
 CC molecule into distinct functional domains. For comparison FimA and
 CC the minor components FimF and FimG only have 2 cysteine residues.
 CC The localisation of the cysteine residues in FimH points to a tandem
 CC arrangement of two ancestral genes. Similar amino acids can be
 CC found in similar positions in the two halves of the FimH protein. The
 CC "midway" point is located roughly around residue 150 in the mature
 CC protein. The two halves or domains of FimH have evolved differently
 CC with the N-terminal section becoming the domain harbouring the receptor
 CC binding site, whereas the C-terminal sector became the domain of the
 CC molecule required for integration into the fimbrial organelle. These
 CC sequences may be used in the production of a variant FimH adhesin which
 CC may be useful for targeting active compounds and microbial cells to
 CC locations comprising selected receptors to which the adhesins bind.
 CC
 XX
 SQ Sequence 300 AA;
 Query Match 97.3%; Score 825; DB 16; Length 300;
 Best Local Similarity 96.9%; Pred. No. 7.2e-81;
 Matches 156; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 PVAVNGNLVVDLSTQIFCHNDYPERITTDVYTLQRGSAVGGVSNFGTVKYGSSVPP 60
 DB 47 PAVAVNGNLVVDLSTQIFCHNDYPERITTDVYTLQRGSAVGGVSNFGTVKYGSSVPP 106
 QY 61 TTSETPRVYNSRTDKPMPALYLTVPVSSAGLVKAGSLIAVLILQTNNNYSDDFQFV 120
 DB 107 TTSETPRVYNSRTDKPMPALYLTVPVSSAGLVKAGSLIAVLILQTNNNYSDDFQFV 166
 QY 121 WNIYANDVYVPTGGCDVSAHDVTVTLDPYRGSVPIPLTVY 161
 DB 167 WNIYANDVYVPTGGCDVSAHDVTVTLDPYRGSVPIPLTVY 207
 RESULT 35
 AAR76772
 ID AAR76772 standard; protein; 300 AA.
 AC AAR76772;
 XX
 DT 15-MAR-1996 (first entry)
 XX
 DE FimH protein derived from E. coli clinical isolate U221-3.

```

XX  FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KM  FimF; FimF; receptor binding site.
XX
XX  Escherichia coli clinical isolate U221-3.
OS
XX
FH  Key
FH  Peptide      Location/Qualifiers
FT  1..21
FT  /note= "Signal peptide"
FT  Protein      22..300
FT  /note= "Mature FimH"
XX
PN  WO9520657-A1.
XX
PD  03-AUG-1995.
XX
XX  27-JAN-1995; 95WO-DK00042.
XX
XX  27-JAN-1994; 94US-0187166.
XX
XX  (GXBI-) GX BIOSYSTEMS AS.
XX
PI  Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX  WPI; 1995-275442/36.
DR  N-PSDB; AA093074.
XX
XX  Receptor specific bacterial adhesins - useful for targetting active
PT  compounds and microbial cells to locations of receptors
XX
XX  Example 1; Page 44-45; 152pp; English.
XX
XX  The sequences given in AAR76763-76 are FimH proteins from various E.
CC  coli clinical isolates. FimH is located at the tip of the type 1
CC  fimbriae and also intercalated at intervals in the fimbrial organelle.
CC  Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC  structures containing terminally located alpha-D-mannoside residues.
CC  FimH contains 4 cysteine residues assumed to direct folding of the
CC  molecule into distinct functional domains. For comparison FimA and
CC  the minor components FimF and FimG only have 2 cysteine residues.
CC  The localisation of the cysteine residues in FimH points to a tandem
CC  arrangement of two ancestral genes. Similar amino acids can be
CC  found in similar positions in the two halves of the FimH protein. The
CC  "midway" point is located roughly around residue 150 in the mature
CC  protein. The two halves or domains of FimH have evolved differently
CC  with the N-terminal section becoming the domain harbouring the receptor
CC  binding site, whereas the C-terminal sector became the domain of the
CC  molecule required for integration into the fimbrial organelle. These
CC  sequences may be used in the production of a variant FimH adhesin which
CC  may be useful for targetting active compounds and microbial cells to
CC  locations comprising selected receptors to which the adhesins bind.
XX
XX  Sequence 300 AA;
SQ

```

Query Match 97.2%; Score 824; DB 16; Length 300;
Best Local Similarity 96.9%; Pred. No. 9.2e-81;
Matches 156; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

QY  1 PAVNVGQNLVVDLSTQIFCHNDYPERITTDYVTLQSGAVGCVLSNFGSYVYKSGSSYPP 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  47 PAVNVGQNLVVDLSTQIFCHNDYPERITTDYVTLQSGAVGCVLSNFGSYVYKSGSSYPP 106
QY  61 TTSETPRVYVNSRTDKPMPALYLTLPVSSAGLVITKAGSLIAVLILRQTNVNSDDFOFV 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  107 TTSETPRVYVNSRTDKPMPALYLTLPVSSAGLVITKAGSLIAVLILRQTNVNSDDFOFV 166
QY  121 WNIYANDVYVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  167 WNIYANDVYVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 207

```

```

ID  AAE18425 standard; Protein; 279 AA.
XX
XX  AAE18425;
AC
XX
XX  07-MAY-2002 (first entry)
DT
XX
XX  Escherichia coli strain EC58 FimH protein.
DE
XX
XX  FimH; immune response; antibacterial; enterobacillus-related disease;
KM  therapy; vaccine; urinary tract infection; bladder.
XX
XX  Escherichia coli EC58.
OS
XX
FH  Key
FH  Location/Qualifiers
FT  Misc-difference 176
FT  /note= "Encoded by COT"
FT  Misc-difference 201
FT  /note= "Encoded by ACC"
XX
XX  WO200204496-A2.
XX
XX  17-JAN-2002.
XX
XX  06-JUL-2001; 2001WO-US21525.
XX
XX  07-JUL-2000; 2000US-216750P.
XX
XX  (MEDI-) MEDIMUNE INC.
XX
XX  Langermann S, Revel A, Auguste C, Burteln J;
PI  WPI; 2002-171702/22.
XX  N-PSDB; AAD29364.
XX
XX  New immunogenic polypeptide, useful as vaccine for protecting against
PT  an enterobacillus-related disease in a patient at risk of contracting
PT  such disease, e.g. urinary tract infection or a bladder infection -
XX
XX  Claim 3; Fig 2; 101pp; English.
XX
XX  The invention relates to bacterial immunogenic agents for administration
CC  to humans and non-human animals to stimulate an immune response. The
CC  invention also relates to methods for vaccination of mammalian species
CC  with variants of E. coli FimH protein derived from different strains of
CC  E. coli. The vaccine composition or the antibody is useful for protecting
CC  against and treating an enterobacillus-related disease in a patient
CC  afflicted or at a risk of contracting the disease. In particular, the
CC  disease is a urinary tract or bladder infection. The disease is caused
CC  by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC  The present sequence is Escherichia coli strain EC58 FimH protein.
XX
XX  Sequence 279 AA;
SQ

```

Query Match 97.1%; Score 823; DB 23; Length 279;
Best Local Similarity 96.3%; Pred. No. 1.1e-80;
Matches 155; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

QY  1 PAVNVGQNLVVDLSTQIFCHNDYPERITTDYVTLQSGAVGCVLSNFGSYVYKSGSSYPP 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  26 PAVNVGQNLVVDLSTQIFCHNDYPERITTDYVTLQSGAVGCVLSNFGSYVYKSGSSYPP 85
QY  61 TTSETPRVYVNSRTDKPMPALYLTLPVSSAGLVITKAGSLIAVLILRQTNVNSDDFOFV 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  86 TTSETPRVYVNSRTDKPMPALYLTLPVSSAGLVITKAGSLIAVLILRQTNVNSDDFOFV 145
QY  121 WNIYANDVYVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  146 WNIYANDVYVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186

```

RESULT 36
AAE18425

RESULT 37
AAR76770
ID AAR76770 standard; protein; 300 AA.

XX AC AAR6770;
XX DT 15-MAR-1996 (first entry)
XX DE FimH protein derived from E. coli clinical isolate F-18.
XX DE FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
XX DE FimH; FimH; receptor binding site.
XX OS Escherichia coli clinical isolate F-18.
XX FH Key
XX FT Peptide 1..21
XX FT /note= "Signal peptide"
XX FT Protein 22..300
XX FT /note= "Mature FimH"
XX PN MO9520657-A1.
XX PD 03-AUG-1995.
XX PF 27-JAN-1995; 95WO-DK00042.
XX PR 27-JAN-1994; 94US-0187166.
XX PA (GXBI-) GX BIOSYSTEMS AS.
XX PI Hasty DL, Klemm P, Molin S, Palsen L, Sokurenko EV;
XX DR WPI: 1995-275442/36.
XX DR N-PSDB: AAO93063.
XX PT Receptor specific bacterial adhesins - useful for targeting active
XX PT compounds and microbial cells to locations of receptors
XX PS Example 1; Page 44-45; 152pp: English.
XX CC The sequences given in AAR6763-76 are FimH proteins from various E.
XX CC coli clinical isolates. FimH is located at the tip of the type 1
XX CC fimbriae and also intercalated at intervals in the fimbrial organelle.
XX CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
XX CC structures containing terminally located alpha-D-mannoside residues.
XX CC FimH contains 4 cysteine residues assumed to direct folding of the
XX CC molecule into distinct functional domains. For comparison FimA and
XX CC the minor components FimF and FimG only have 2 cysteine residues.
XX CC The localization of the cysteine residues in FimH points to a tandem
XX CC arrangement of two ancestral genes. Similar amino acids can be
XX CC found in similar positions in the two halves of the FimH protein. The
XX CC "midway" point is located roughly around residue 150 in the mature
XX CC protein. The two halves or domains of FimH have evolved differently
XX CC with the N-terminal section becoming the domain harbouring the receptor
XX CC binding site, whereas the C-terminal sector became the domain of the
XX CC molecule required for integration into the fimbrial organelle. These
XX CC sequences may be used in the production of a variant FimH adhesin which
XX CC may be useful for targeting active compounds and microbial cells to
XX CC locations comprising selected receptors to which the adhesins bind.
XX SQ Sequence 300 AA;
XX
Query Match 96.9%; Score 822; DB 16; Length 300;
Best Local Similarity 96.3%; Pred. No. 1.5e-80;
Matches 155; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB 167 WNIYANDVVPPTGCGDVASRDVTVTLPDYRGVPIPLTVY 207
|||||
RESULT 38
AAE18414
ID AAE18414 standard; Protein; 279 AA.
XX AC AAE18414;
XX DT 07-MAY-2002 (first entry)
XX DE Escherichia coli strain B212 FimH protein.
XX DE FimH; immune response; antibacterial; enterobacillus-related disease;
XX DE therapy; vaccine; urinary tract infection; bladder.
XX OS Escherichia coli B212.
XX FH Key
XX FT Misc-difference 176
XX FT /note= "Encoded by COT"
XX FT Misc-difference 201
XX FT /note= "Encoded by ACC"
XX PN MO200204496-A2.
XX PD 17-JAN-2002.
XX PF 06-JUL-2001; 2001WO-US21525.
XX PR 07-JUL-2000; 2000US-216750P.
XX PA (MEDI-) MEDIMUNE INC.
XX PI Langermann S, Revel A, Auguste C, Burteln J;
XX DR WPI: 2002-171702/22.
XX DR N-PSDB: AAD29353.
XX PT New immunogenic polypeptide, useful as vaccine for protecting against
XX PT an enterobacillus-related disease in a patient at risk of contracting
XX PT such disease, e.g. urinary tract infection or a bladder infection
XX PS Claim 3; Fig 2; 101pp: English.
XX CC The invention relates to bacterial immunogenic agents for administration
XX CC to humans and non-human animals to stimulate an immune response. The
XX CC invention also relates to methods for vaccination of mammalian species
XX CC with variants of E. coli FimH protein derived from different strains of
XX CC E. coli. The vaccine composition or the antibody is useful for protecting
XX CC against and treating an enterobacillus-related disease in a patient
XX CC afflicted or at a risk of contracting the disease. In particular, the
XX CC disease is a urinary tract or bladder infection. The disease is caused
XX CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
XX CC The present sequence is Escherichia coli strain B212 FimH protein.
XX SQ Sequence 279 AA;
XX
Query Match 96.6%; Score 819; DB 23; Length 279;
Best Local Similarity 95.7%; Pred. No. 2.9e-80;
Matches 154; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db	146	WNIIYANDVVVPTGGCDASARDVTTLDPDYGSPVIPLTVY	186
<hr/>			
RESULT 39			
ID	AAE18423	standard; Protein; 279 AA.	
XX	AAE18423;		
AC			
XX			
DT	07-MAY-2002	(first entry)	
XX			
DE	Escherichia coli strain EC45 FimH protein.		
KM	FimH; Immune response; antibacterial; enterobacillus-related disease;		
XW	therapy; vaccine; urinary tract infection; bladder.		
XX			
OS	Escherichia coli EC45.		
XX			
PN	WO200204496-A2.		
PJ	17-JAN-2002.		
PD			
XX			
PF	06-JUL-2001; 2001WO-US21525.		
XX			
PR	07-JUL-2000; 2000US-216750P.		
PA	(MED)- MEDIMUNE INC.		
PI	Langermann S, Revel A, Auguste C, Burlin J;		
PT	WPI: 2002-171702/22.		
DR	N-PsDB: AAD29362.		
XX			
PS	New immunogenic polypeptide, useful as vaccine for protecting against		
XX	an enterobacillus-related disease in a patient at risk of contracting		
XX	such disease, e.g. urinary tract infection or a bladder infection -		
XX	Claim 3; Fig 2; 10pp; English.		
CC	The invention relates to bacterial immunogenic agents for administration		
CC	to humans and non-human animals to stimulate an immune response. The		
CC	invention also relates to methods for vaccination of mammalian species		
CC	with variants of E. coli FimH protein derived from different strains of		
CC	E. coli. The vaccine composition or the antibody is useful for protecting		
CC	against and treating an enterobacillus-related disease in a patient		
CC	afflicted or at a risk of contracting the disease. In particular, the		
CC	disease is a urinary tract or bladder infection. The disease is caused		
CC	by a bacterium of the family Enterobiaceae, particularly E. coli.		
CC	The present sequence is Escherichia coli strain EC45 FimH protein.		
XX			
SQ	Sequence 279 AA:		
<hr/>			
Query Match	96.6%; Score 819; DB 23; Length 279;		
Best Local Similarity	95.7%; Pred. No. 2.9e-80;		
Matches 154; Conservative 4; Mismatches 3; Indels 0; Gaps 0;			
QY	1 PYYNNGONLVLDLSTOIFCHNDYPETITDYVLQGSATGGLSNFSGTYKXSSGYPP	60	
DB	26 PAVNWGNLVLDLSTOIFCHNDYPETITDYVLQGAAGVLSFSFGTYKNSSPYPP	85	
QY	61 TTSETPRVYNSTRTPMPVALYLTPVSSAGGLVKAGSLIIVLLIRONTNNNSDDFCFY	120	
DB	86 TTSERPRVYNSTRTPMPVALYLTPVSSAGGVAIKAGSLIIVLLIRONTNNNSDDFCFY	145	
QY	121 WNIYANDVVVPTGGCDVSARVDYTTLDPYRGSVPILTVY	161	
DB	146 WNIYANDVVVPTGGCDVSARVDYTTLDPYRGSVPILTVY	186	
<hr/>			
RESULT 40			
ID	AAE18431	standard; Protein; 279 AA.	
XX	AAE18431		

AC	AAE18431:					
XX						
DT	07-MAY-2002 (first entry)					
XX						
DE	Escherichia coli strain EC95 FimH protein.					
KW	FimH; immune response; antibacterial; enterobacillus-related disease;					
KM	therapy; vaccine; urinary tract infection; bladder.					
XX						
OS	Escherichia coli EC95.					
FH						
FT	Key	Location/Qualifiers				
FT	Misc-difference 176	/note= "Encoded by CCR"				
FT	Misc-difference 201	/note= "Encoded by ACC"				
XX						
FN	WO200204496-A2.					
PD	17-JAN-2002.					
XX						
PF	06-JUL-2001; 2001WO-US21525.					
XX						
PR	07-JUL-2000; 2000US-216750P.					
XX						
PA	(MED1-) MEDIMUNE INC.					
XX						
PI	Langermann S, Revel A, Auguste C, Burlin J;					
DR	WPI: 2002-171702/22.					
XX	N-PsDB: AAD29370.					
PT						
PT	New immunogenic polypeptide, useful as vaccine for protecting against					
PT	an enterobacillus-related disease in a patient at risk of contracting					
PT	such disease, e.g. urinary tract infection or a bladder infection -					
XX						
XX	Claim 3; Page 87-88; 101pp; English.					
XX						
CC	The invention relates to bacterial immunogenic agents for administration					
CC	to humans and non-human animals to stimulate an immune response. The					
CC	invention also relates to methods for vaccination of mammalian species					
CC	with variants of E. coli FimH protein derived from different strains of					
CC	E. coli. The vaccine composition or the antibody is useful for protecting					
CC	against and treating an enterobacillus-related disease in a patient					
CC	afflicted or at a risk of contracting the disease. In particular, the					
CC	disease is a urinary tract or bladder infection. The disease is caused					
CC	by a bacterium of the family Enterobacteriaceae, particularly E. coli.					
CC	The present sequence is Escherichia coli strain EC95 FimH protein.					
XX						
SQ	Sequence 279 AA;					
<hr/>						
Query Match	96.6%;	Score 819;	DB 23;	Length 279;		
Best Local Similarity	96.3%;	Pred. No. 2.9e-80;				
Matches 155;	Conservative 3;	Mismatches 3;	Indels 0;	Gaps 0;		
OY	1 PAVVNGONLVVDLSIQIFCHNDYPETITPYVLORSAAGVLSNPSGTVKXSGSSPPP 60					
Db	:					
	26 PAVVNGONLVVDLSIQIFCHNDYPETITPYVLORSAAGVLSNPSGTVKXSGSSPPP 85					
OY	61 TTSETTPRVVYNSRTDKPMWVALYLTPVSSAGGLVKAGSLIAVLILRQTNNYNSSDDFOFV 120					
Db	:					
	86 TTSETTPRVVYNSRTDKPMWVALYLTPVSSAGGLVKAGSLIAVLILRQTNNYNSSDDFOFV 145					
OY	121 WNIYANDVVYPTGGCDVSARDVYTLPYRKRSVPPLPLY 161					
Db	:					
	146 WNIYANDVVYPTGGCDVSARDVYTLPYRKRSVPPLPLY 186					
<hr/>						
RESULT 41						
AAR676775						
ID	AAR676775 standard; protein; 300 AA.					
AC	AAR676775;					

```

XX 15-MAR-1996 (first entry)
DT
XX
XX FimH protein derived from E. coli clinical isolate MU#11-2.
DE
XX
XX FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW FimH; FimH; receptor binding site.
XX
OS Escherichia coli clinical isolate MU#11-2.
XX
XX
XX Key Location/Qualifiers
FH Peptide 1..21
FT /note="Signal peptide"
FT 22..300
FT Protein /note="Mature FimH"
XX
XX
XX W09520657-A1.
XX
XX
XX 03-AUG-1995.
XX
XX 27-JAN-1995; 95WO-DK00042.
XX
XX 27-JAN-1994; 94US-0187166.
XX
XX (GXB1-) GX BIOSYSTEMS AS.
XX
XX Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
PI
XX MPI: 1995-275442/36.
DR N-PSDB; AAC93064.
XX
XX Receptor specific bacterial adhesins - useful for targeting active
PT compounds and microbial cells to locations of receptors
XX
XX Example 1; Page 44-45; 152pp; English.
XX
XX The sequences given in AAR76763-76 are FimH proteins from various E.
CC coli clinical isolates. FimH is located at the tip of the type 1
CC fimbriae and also intercalated at intervals in the fimbrial organelle.
CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC structures containing terminally located alpha-D-mannoside residues.
CC FimH contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison FimA and
CC the minor components FimF and FimG only have 2 cysteine residues.
CC The localisation of the cysteine residues in FimH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FimH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FimH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of a variant FimH adhesin which
CC may be useful for targeting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.
XX
XX
XX Sequence 300 AA:
SQ
Query Match 96.5%; Score 818; DB 16; Length 300;
Best Local Similarity 95.7%; Pred. No. 4.1e-80;
Matches 154; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

RESULT 42
ABB09458
ID ABB09458 standard; Protein; 408 AA.
XX
XX ABB09458;
AC
XX
XX 01-JUL-2002 (first entry)
DT
XX
XX Fusion protein prty-FimH-prty.
DE
XX
XX Exoproteinase; immunostimulant; vaccine; anchor peptide;
KW FimH; ciliated adhesive factor.
XX
XX Lactobacillus helveticus.
OS
OS Escherichia coli.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Peptide 1..70
FT /label="signal_peptide"
FT /note="derived from L. helveticus"
FT Protein 71..295
FT /label="FimH_fragment"
FT /note="derived from E. coli ciliated adhesive factor"
FT Peptide 296..408
FT /label="anchor_peptide"
FT /note="derived from L. helveticus"
XX
XX JP2002017357-A.
XX
XX 22-JAN-2002.
XX
XX 04-JUL-2000; 2000JP-0202442.
XX
XX 04-JUL-2000; 2000JP-0202442.
XX
XX (CALV ) CALPIS SHOKUHIN KOGYO KK.
XX
XX MPI; 2002-221706/28.
DR N-PSDB; ABL52756.
XX
XX New anchor peptide, useful for anchoring protein onto microbe -
PT
XX
XX Example 1; Page 10-11; 12pp; Japanese.
XX
XX The invention relates to an anchor peptide derived from the
XX exoproteinase of Lactobacillus helveticus FERM BP-6060.
XX CC The activity of compositions of the invention may be described as
XX immunostimulatory. The invention also includes a fusion protein,
XX containing the anchor peptide and a signal peptide fused to a useful
XX protein. The invention also includes a method for anchoring a useful
XX protein onto the surface of a microbe. The microbe can be used as a
XX vaccine. The current sequence represents a fusion protein referred to as
XX prty-FimH-prty. This fusion protein contains a fragment of the FimH
XX protein, which is a ciliated adhesive factor from E. coli, arranged
XX between the signal sequence and anchor sequences derived from L.
XX helveticus exoproteinase (see ABB09456 and ABB09457).
XX
XX
XX Sequence 408 AA:
SQ
Query Match 96.2%; Score 816; DB 23; Length 408;
Best Local Similarity 95.7%; Pred. No. 1.1e-79;
Matches 154; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```


XX Sequence 300 AA;
 SQ
 Query Match 96.0%; Score 814; DB 16; Length 300;
 Best Local Similarity 95.7%; Pred. No. 1.1e-79;
 Matches 154; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 PAVNWGQNLVVDLSTQIFCHNDYPTITDVTYLRGSAVGGVLSFSGTVKYNSSYPP 60
 DB 47 PAVNWGQNLVVDLSTQIFCHNDYPTITDVTYLRGSAVGGVLSFSGTVKYNSSYPP 106
 QY 61 TTSETPRVYNSRTDKPMPALYLPVSSAGLVKAGSLIAVLILRQTNNNNSDDFOFV 120
 DB 107 TTSETPRVYNSRTDKPMPALYLPVSSAGLVKAGSLIAVLILRQTNNNNSDDFOFV 166
 QY 121 WNIYANDVYVPTGGCDVSARDVTYTLDPYRGSVPIPLTVY 161
 DB 167 WNIYANDVYVPTGGCDVSARDVTYTLDPYRGSVPIPLTVY 207

RESULT 45

AA076764
 ID AAR76764 standard; protein; 300 AA.

AA076764;
 AC
 XX
 XX
 DT
 XX

15-MAR-1996 (first entry)

FimH protein derived from E. coli clinical isolate CI#12.

FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
 FimA; FimF; FimG; receptor binding site.

Escherichia coli clinical isolate CI#12.

Key Location/Qualifiers
 FT Peptide 1..21
 FT /note= "Signal peptide"
 FT Protein 22..300
 FT /note= "Mature FimH"

W09520657-A1.

03-AUG-1995.

27-JAN-1995; 95WO-DK00042.

27-JAN-1994; 94US-0187166.

(GXBI-) GX BIOSYSTEMS AS.

Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;

WPI: 1995-275442/36.

N-PSDB: AA093075.

Receptor specific bacterial adhesins - useful for targetting active
 compounds and microbial cells to locations of receptors

Example 1; Page 44-45; 152pp; English.

The sequences given in AAR76763-76 are FimH proteins from various E.
 coli clinical isolates. FimH is located at the tip of the type 1
 fimbriae and also intercalated at intervals in the fimbrial organelle.
 Most forms of the FimH adhesin target to, and bind to, oligosaccharide
 structures containing terminally located alpha-D-mannoside residues.
 FimH contains 4 cysteine residues assumed to direct folding of the
 molecule into distinct functional domains. For comparison FimA and
 the minor components FimF and FimG only have 2 cysteine residues.
 The localization of the cysteine residues in FimH points to a tandem
 arrangement of two ancestral genes. Similar amino acids can be
 found in similar positions in the two halves of the FimH protein. The
 "midway" point is located roughly around residue 150 in the mature

CC protein. The two halves or domains of FimH have evolved differently
 CC with the N-terminal section becoming the domain harbouring the receptor
 CC binding site, whereas the C-terminal sector became the domain of the
 CC molecule required for integration into the fimbrial organelle. These
 CC sequences may be used in the production of a variant FimH adhesin which
 CC may be useful for targetting active compounds and microbial cells to
 CC locations comprising selected receptors to which the adhesins bind.
 XX

Sequence 300 AA;

Query Match 96.0%; Score 814; DB 16; Length 300;
 Best Local Similarity 95.7%; Pred. No. 1.1e-79;
 Matches 154; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAVNWGQNLVVDLSTQIFCHNDYPTITDVTYLRGSAVGGVLSFSGTVKYNSSYPP 60
 DB 47 PAVNWGQNLVVDLSTQIFCHNDYPTITDVTYLRGSAVGGVLSFSGTVKYNSSYPP 106
 QY 61 TTSETPRVYNSRTDKPMPALYLPVSSAGLVKAGSLIAVLILRQTNNNNSDDFOFV 120
 DB 107 TTSETPRVYNSRTDKPMPALYLPVSSAGLVKAGSLIAVLILRQTNNNNSDDFOFV 166
 QY 121 WNIYANDVYVPTGGCDVSARDVTYTLDPYRGSVPIPLTVY 161
 DB 167 WNIYANDVYVPTGGCDVSARDVTYTLDPYRGSVPIPLTVY 207

Search completed: November 28, 2002, 17:19:55
 Job time : 68 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 10:36:15 ; Search time 24 Seconds

(without alignments)
278.237 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186

Perfect score: 848
Sequence: 1 PVTN/GONLVVDLSTQIFCH.....DVTVLDPYRGSVPIPLFVY 161

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	83.4	98.3	300	1	FIMH_ECOLI
2	38.4	45.3	304	1	YDEQ_ECOLI
3	205	24.2	291	1	SFAH_ECOLI
4	86.5	10.2	550	1	HEMA_IADH6
5	85.5	10.1	315	1	HEMA_VACCI
6	85.5	10.1	550	1	HEMA_IADH3
7	84.5	10.0	566	1	HEMA_IADCO
8	83.5	9.8	567	1	HEMA_IADVI7
9	83.5	9.8	639	1	APPR_ENTH1
10	83	9.8	519	1	LAC2_TRAVI
11	82.5	9.7	315	1	HEMA_VACCT
12	81.5	9.6	550	1	HEMA_IADH2
13	80.5	9.5	550	1	HEMA_IADH2
14	80	9.4	314	1	HEMA_IADH2
15	80	9.4	385	1	HEMA_VACCV
16	79	9.3	373	1	NK3R_MOUSE
17	78.5	9.3	550	1	PLYD_ASPPG
18	78.5	9.3	550	1	HEMA_IADH5
19	77.5	9.1	249	1	HEMA_IADH7
20	77.5	9.1	315	1	HEMA_IADH3
21	77.5	9.1	550	1	HEMA_VACCC
22	77.5	9.1	550	1	HEMA_IADH1
23	77.5	9.1	550	1	HEMA_IADH4
24	77.5	9.1	550	1	HEMA_IADH4
25	77.5	9.1	550	1	HEMA_IADH3
26	76.5	9.0	550	1	HEMA_IADH3
27	76.5	9.0	550	1	HEMA_IADH3
28	76.5	9.0	550	1	HEMA_IADH3
29	76	9.0	371	1	YADC_SCHPO
30	76	9.0	558	1	LYVD_SULO
31	76	9.0	928	1	PH11_CHLPP
32	75.5	8.9	566	1	HEMA_IADH3
33	75.5	8.9	566	1	HEMA_IADH3

34	75	8.8	519	1	LAC2_TRAVE	012718	trametes ve
35	74.5	8.8	375	1	PEUC_ERWCH	P11073	erwinia chr
36	74.5	8.8	566	1	HEMA_IADH3	P03442	influenza a
37	74.5	8.8	2195	1	POUG_ECLIG	P29813	e genome po
38	73.5	8.7	328	1	HEMA_IADH6	P04664	influenza a
39	73	8.6	452	1	NK3R_RAT	P16177	rattus norv
40	73	8.6	591	1	FLGE_CAVCR	P35806	caulobacter
41	73	8.6	1656	1	OMPB_RICJA	006653	r outer mem
42	72.5	8.5	383	1	GAB2_HUMAN	006547	homo sapien
43	72.5	8.5	580	1	FLA2_PYRKO	09v20	pyrococcus
44	72	8.5	342	1	RUVA_RICPR	09v20	pyrococcus
45	72	8.5	550	1	HEMA_IADH4	P43259	influenza a

ALIGNMENTS

RESULT 1
FIMH_ECOLI
STANDARD: PRT: 300 AA.
DIAGRAM: 01-FEB-1995 (Rel. 31, last sequence update)
16-OCT-2001 (Rel. 40, last annotation update)
FIMH protein precursor.
FIMH OR B4320.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88038337; PubMed=2890081;
RA Klemm P., Christensen G.;
RT "Three fim genes required for the regulation of length and mediation
of adhesion of Escherichia coli type 1 fimbriae."
RL Mol. Gen. Genet. 208:439-445(1987).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burdland V.D., Plunkett G., III, Sofia H.J., Daniels D.L.,
Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
region from 92.8 through 100 minutes."
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [13]
RP CHARACTERIZATION.
RX MEDLINE=90256291; PubMed=1971261;
RA Krogfelt K.A., Bergmans H., Klemm P.;
RT "Direct evidence that the fimH protein is the mannose-specific
adhesin of Escherichia coli type 1 fimbriae."
RL Infect. Immun. 58:1995-1998(1990).
CC -I- FUNCTION: INVOLVED IN REGULATION OF LENGTH AND MEDIATION OF
ADHESION OF TYPE 1 FIMBRIAE (BUT NOT NECESSARY FOR THE PRODUCTION
OF FIMBRIAE). ADHESIN RESPONSIBLE FOR THE BINDING TO D-MANNOSE. IT
IS LATERALLY POSITIONED AT INTERVALS IN THE STRUCTURE OF THE TYPE
1 FIMBRIAE. IN ORDER TO INTEGRATE FIMH IN THE FIMBRIAE FIMF AND
FIMG ARE NEEDED.
CC -I- SIMILARITY: THIS PROTEIN EXHIBITS HOMOLOGY WITH THE FIMBRIAL
SUBUNIT PROTEIN FIMH.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC EMBL: X05672; CAA29156.1; -
DR EMBL: U14003; AAA97216.1; -
DR EMBL: AE000502; AAC77276.1; -

```

DR      PIR: S09563; S09563.
DR      Ecogeene: EG10315; famH.
DR      InterPro: IPR000259; Fimbrial.
DR      Pfam: PF00419; Fimbrial; 1.
DR      Fimbrila; Signal; Complete proteome.
KT      SIGNAL 1 23 POTENTIAL.
FT      CHAIN 24 300 FIMB PROTEIN.
FT      CONFLICT 197 197 P -> R (IN REF. 1).
FT      CONFLICT 222 222 T -> H (IN REF. 1).
SO      SEQUENCE 300 AA; 31473 MW; 939204AA51658747D CMC64;

Query Match 98.3%; Score 834; DB 1; Length 300;
Best Local Similarity 98.1%; Pred. No. 1,3e-70;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY      1 PVPVNGQLVYDLSIQICFCHNDYPTETIDYVTLONGSAVGGVLSNFSGTVKSGSSYFP 60
DB      47 PVPVNGQLVYDLSIQICFCHNDYPTETIDYVTLONGSAVGGVLSNFSGTVKSGSSYFP 106
OY      61 TTSETPRVVNSRTDKPPVALYLPVSSAGGLYTKAGSLAVLLRQTNNNYSDDFQEV 120
DB      107 TTSETPRVVNSRTDKPPVALYLPVSSAGGVALTKAGSLAVLLRQTNNNYSDDFQEV 166
OY      121 WNIYANDVYVPTGGCDVSARDYVTLTPDYKRSVPIPLTV 161
DB      167 WNIYANDVYVPTGGCDVSARDVYTLTPDYPSGVPIPLTV 207

RESULT 2
YDEQ_ECOLI
ID      YDEQ_ECOLI STANDARD: PRT: 304 AA.
AC      P77588;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical fimbrial-like protein ydeq precursor.
GN      Ydeq OR B1502.
OC      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=97426617; PubMed=9278503;
RX      Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      "The complete genome sequence of Escherichia coli K-12.";
RL      Science 277:1453-1474(1997).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12;
RX      MEDLINE=97251357; PubMed=9097039;
RA      Alha H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA      Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA      Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA      Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA      Tampei G., Seli Y., Sivasundaram S., Tagami H., Takeda J.,
RA      Takekoshi K., Takeuchi Y., Wada C., Yamamoto Y., Horichi T.;
RT      "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL      DNA Res. 3:363-377(1996).
-----
-i SIMILARITY: TO TYPE-1 FIMBRIAL SUBUNITS.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
```

[illegible]

```

RESULT 3
SFAM_ECOLI
ID SFAM_ECOLI STANDARD: PRT: 291 AA.
AC PI3431.
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE S-fimbrial adhesin protein sfah precursor.
GN SFAM.
OS Escherichia coli.
OC Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:K15;
RX MEDLINE=90158121; PubMed=2576095;
RA Schmolli T., Hoshuetzky H., Morschhauser J., Lottspeich F.,
RA Jann K., Hacker J.;
RT "Analysis of genes coding for the stialic acid-binding adhesin and two
RT other minor fimbrial subunits of the S-fimbrial adhesin determinant
RT of Escherichia coli. ";
RL Mol. Microbiol. 3:1735-1744(1989).
CC -I- FUNCTION: FIMBRAE (ALSO CALLED PILI), POLAR FILAMENTS RADIATING
CC FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5
CC MICROMETERS AND NUMBERING 100-300 PER CELL. ENABLE BACTERIA TO
CC COLONIZE THE EPITHELIUM OR SPECIFIC HOST ORGANS.
CC -I- FUNCTION: THIS PROTEIN HAS SOME INFLUENCE ON THE LEVEL OF
CC FIMBRINATION OF THE CELL.
CC -I- DISEASE: S-FIMBRIAL ADHESINS ENABLE PATHOGENIC E. COLI CAUSING
CC URINARY-TRACT INFECTIONS OR NEWBORN MENINGITIS TO ATTACH TO
CC GLYCOPROTEINS TERMINATING WITH ALPHA-SIALIC ACID-(2-3)-BETA-GAL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X16664; CAA34654.1; -.
CC PIR: S15927; S15927.
CC InterPro: IPR000259; Fimbrial.

```

DR Pfam: PF00419; Fimbrin; 1.
 KW Fimbrin; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 291
 SQ SEQUENCE 291 AA; 31297 MW; E026AA4713C2F947 CRC64;
 Query Match 24.2%; Score 205; DB 1; Length 291;
 Best Local Similarity 32.6%; Pred. No. 5.6e-12;
 Matches 56; Conservative 22; Mismatches 82; Indels 12; Gaps 5;
 QY 1 PVTNAGQNL-VVDLSTQIFCHN-DYPTITDYTLQSGAVGVL-----NFGTVK 51
 DB 34 PVEYTKSIVLDSQVSCQNDSTGQNTDYLIKSGSFALDTKTYGRDFTSRPT 93
 QY 52 YGSSYPTPTSETPRVYNSRTDKPMPVALYLPVSSAGLVKAGSLIALLRQ--T 109
 DB 94 GYARQLPQLGDFLQVTEAFYGYGWMKPPAKLYLPERGVGKVIYNGDLATLYVKFST 153
 QY 110 NNNNSDDEPQVNNIYANDVYVPTGGCDVSARDVTLTPRGRS-VPIPLTV 160
 DB 154 KGOAGERNPTMRFYATNDVHIQTGTCRVSSNNVKYDLFPYGGPVTVPLTV 205
 RESULT 4
 HEMA_IADH6 STANDARD; PRT; 550 AA.
 ID HEMA_IADH6
 AC P12587; Q84017;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
 GN HA.
 OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 NC NCBL_TaxID=11362;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87265458; Pubmed=2440178;
 RA Kida H., Kawasaka Y., Naeve C.W., Webster R.G.;
 RT "Antigenic and genetic conservation of H3 influenza virus in wild
 RT ducks";
 RT Virology 159:109-119(1987).
 RL
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M16742; AAA43148.1; -
 DR PIR: F27813; HMTV98.
 DR HSSP: P03437; HMGJ.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR ProDom: PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON_TER 1 1
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 8 8 Y -> N (IN PIR DATA BANK).
 SQ SEQUENCE 550 AA; 61711 MW; 67BC05F44736CFE CRC64;
 Query Match 10.2%; Score 86.5; DB 1; Length 550;
 Best Local Similarity 25.0%; Pred. No. 1.3;
 Matches 37; Conservative 18; Mismatches 58; Indels 35; Gaps 6;
 QY 21 NNYPTITDYVTLQSGANGVL-----NFGTVKXSGSSYPTPTSETPRVYNSRT 74
 DB 96 NCYPYDVPDYASRLSVASSGTLFETEGFTWTGYTONGNSN-----ACKRGPSGFFSRLL 151
 QY 75 DKPMPVALYLPVSSAGLVKAGSLIALLRQTNVNSDDEPQVNNIY-----ANN 127
 DB 152 N--W-----LTKSGSTYPLANTMRFYATNDVHIQTGTCRVSSNNVKYDLFPYGGPVTVPLTV 193
 QY 128 DVVPTGGCDVSARDVTLTPRGRSVP 155
 DB 194 LVYQASGVTVSTRSQOTIIPNIGSRP 221
 RESULT 5
 HEMA_VACCI STANDARD; PRT; 315 AA.
 ID HEMA_VACCI
 AC P08714;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Hemagglutinin precursor.
 GN HA OR A56R.
 OS Vaccinia virus (strain 1HD-J).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 NC NCBL_TaxID=10251;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86181588; Pubmed=3008418;
 RA Shida H.;
 RT "Nucleotide sequence of the vaccinia virus hemagglutinin gene";
 RT Virology 150:451-462(1986).
 RL
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- PTM: THE MAJOR 85 KDa GLYCOPROTEIN IS PHOSPHORYLATED AND SULFATED.
 CC O-LINKED GLYCOSYLATION IS REQUIRED FOR HEMAGGLUTINATION AND
 CC HEMADSORPTION ACTIVITIES OF INFECTED CELL MEMBRANES.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M14783; AAA48251.1; -
 DR PIR: A27789; HNVZYV.
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig; 1.
 DR SMART: SM00409; Ig; 1.
 KW Hemagglutinin; Envelope protein; Late protein; Glycoprotein; Signal;
 KW Transmembrane; Immunoglobulin domain; Phosphorylation; Sulfation.
 FT SIGNAL 1 16
 FT CHAIN 17 315 HEMAGGLUTININ.
 FT DOMAIN 17 279 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 280 303 POTENTIAL.
 FT DOMAIN 304 315 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 317 121 IG-LIKE V-TYPE DOMAIN.
 FT DISULFID 34 103
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

SQ SEQUENCE 315 AA; 34834 MW; 25EE5C7AE6A19779 CRC64;
Query Match 10.1%; Score 85.5; DB 1; Length 315;
Best Local Similarity 25.5%; Pred. No. 0.85;
Matches 36; Conservative 17; Mismatches 37; Indels 51; Gaps 7;
QY 24 PETITDYVTLQKSGAVGVLNFGSTVYK-----SSGSPYPTPTSETPRVYNS 72
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 171 PERIDNV-----EDHTDVTYSDSINTVSASGES---TTDEPPTDK 214
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 73 RDKPMPVALYLPVSSAGGLVYKAGSLIALLRQNNYNSDDQFVNNITANDVYVP 132
   : | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 215 EEDHTVTDVSTVSTYSTRSSGIV-----TKSTYDDAD-LYDVTNDQVTP 259
   : | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 133 T--GGCDVSARDVYTLDPYR 151
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 260 TTVGGS-----TTSISNYK 273
   | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
HEMA_IADH3 STANDARD; PRT; 550 AA.
ID HEMA_IADH3
AC P12584; 084012; 089793;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/33/80).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11359;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Neeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks."
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER, EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: M16739; AAA43145.1; -.
DR PIR: C27813; HMTV33.
DR HSSP: P03437; 2VIU.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT CHAIN 1 1 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61577 MW; 6C30BF67CFDCBDDE CRC64;
Query Match 10.1%; Score 85.5; DB 1; Length 550;

```

```

Best Local Similarity 25.0%; Pred. No. 1.7;
Matches 37; Conservative 18; Mismatches 58; Indels 35; Gaps 6;
QY 21 NDYPTTIDYVTLQKSGAVGVL-----NFGSTVYKSSGSPPTPTSETPRVYNSRT 74
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 96 NCYPADVPDYLKSLVASSGTLFEITGFTWGTQVNGSGN---ACKRQPAAGFFSRL 151
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 75 DKPMPVALYLPVSSAGGLVYKAGSLIALLRQNNYNSDDQFVNNITANDVYVP 127
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 152 N--W-----LTKSGSTYPLVNTMPNNDNFKL-YIWDVHHPTNQDQTN 193
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 128 DVVPTGGCDVSARDVYTLDPYRGSVP 155
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 194 LVYQASGGVTVSTRSQOTITIPNIGSRP 221
   | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
HEMA_IAZCO STANDARD; PRT; 566 AA.
ID HEMA_IAZCO
AC P26139;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Swine/Colorado/1/77).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11493;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114135; PubMed=1731092;
RA Bean W.J., Schell M., Katz J., Kawaoaka Y., Neeve C., Gorman O.,
RA Webster R.G.;
RT "Evolution of the H3 influenza virus hemagglutinin from human and
RT nonhuman hosts."
RL J. Virol. 66:1129-1138(1992).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER, EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: M73774; -. NOT_ANNOTATED_CDS.
DR HSSP: P03437; 2VIU.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63295 MW; 07297DB3197C4E5 CRC64;
Query Match 10.0%; Score 84.5; DB 1; Length 566;

```

Best Local Similarity 26.4%; Pred. No. 2.1;
Matches 39; Conservative 18; Mismatches 56; Indels 35; Gaps 8;

```

RESURT 8
HEMA_IATV7 STANDARD: PRT: 567 AA.
AC P03435;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HAI chain;
DE Hemagglutinin HAZ chain].
GN HA.
OS Influenza A virus (strain A/Victoria/3/75).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11483;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80155186; PubMed=6153930;
RA Min Jou W., Verhoeven M., Devos R.,
RA Huylebroeck D., Fiers W., Threlfall G., Barber C., Carey N.,
RA Emtage S.;
RT "Complete structure of the hemagglutinin gene from the human
RT Influenza A/Victoria/3/75 (H3N2) strain as determined from cloned
RT DNA.";
RL Cell 19:683-696(1980).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=80254693; PubMed=7402351;
RA Verhoeven M., Fang R., Jou W.M., Devos R., Huylebroeck D., Saman E.,
RA Fiers W.;
RT "Antigenic drift between the haemagglutinin of the Hong Kong
RT Influenza strains A/Alchi/2/68 and A/Victoria/3/75.";
RL Nature 286:771-776(1980).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: V01098; CAA24281.1; -
CC EMBL: V01086; CAA24270.1; -
CC DR EMBL: V01086; CAA24270.1; -
CC DR PIR: A04050; HMTIV.
CC DR HSSP: P03437; 2VIO.
CC InterPro: IPR001364; Hemagglutn.
CC Pfam: PF00509; Hemagglutinin; 1.
CC PRINTS: PR00329; HEMAGGLUTN12.
CC PRODOM: PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; glycoprotein; signal.
KW SIGNAL
FT CHAIN 1 16 HEMAGGLUTININ HAI CHAIN.
FT CHAIN 17 345 HEMAGGLUTININ HAZ CHAIN.

```

FT	CHAIN	347	567	HEMAGGLUTININ HAZ CHAIN.
FT	CARBOHYD	24	24	N-LINKED (GLCMNC. . .) (POTENTIAL).
FT	CARBOHYD	25	25	N-LINKED (GLCMNC. . .) (POTENTIAL).
FT	CARBOHYD	39	39	N-LINKED (GLCMNC. . .) (POTENTIAL).
FT	CARBOHYD	55	55	N-LINKED (GLCMNC. . .) (POTENTIAL).
FT	CARBOHYD	80	80	N-LINKED (GLCMNC. . .) (POTENTIAL).
FT	CARBOHYD	143	143	N-LINKED (GLCMNC. . .) (POTENTIAL).
FT	CARBOHYD	182	182	N-LINKED (GLCMNC. . .) (POTENTIAL).
FT	CARBOHYD	302	302	N-LINKED (GLCMNC. . .) (POTENTIAL).
FT	CARBOHYD	500	500	N-LINKED (GLCMNC. . .) (POTENTIAL).
SQ	SEQUENCE	567 AA;	63422 MW;	8249D98A80BDC5DEF CR664;

Query Match	9.8%;	Score 83.5;	DB 1;	length 567;
Best Local Similarity	26.4%;	Pred. No. 2.6;		
Matches	39;	Conservative 17;	Mismatches 57;	Indels 35;
				Gaps 7;

```

QY      21 NDPEPTTIDYVTLORGASVGGVL-----NFGSTVXSSSSYFPTTSETPRVYNSNR 74
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      113 NCYPADYVDYASLTSRLSVASSGTLEEFINEGFNMTCVTONGS----ACKRGDPGSEPR 168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      75 DKPMVVALYKLPVSSAGSLVYKAGSLAVLLRQTNNVNSDDDFQVWNIY-----ANN 127
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      169 N--W-----LY-----KSGSTYPVQAVNTPMNDNSDKL-YIVGVHPSTDEQTN 210
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      128 DVVVPTGGCDVSARDVYTLDFDNGSV 155
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      211 LVYQASGVYVSTKRSDQTIIPNWSGR 238
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9	
APRP_ENTHI	
ID	APRP_ENTHI
STANDARD:	
PRT:	639 AA

DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic protein p1 (Pathogenic protein 1) (fragment).
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HM-1:IMSS;
RX MEDLINE=8,926955; Pubmed=2544890;
RA Tannich-E., Horstmann R.D., Knobloch J., Arnold H.H.;
RT "Genomic DNA differences between pathogenic and nonpathogenic
RT Entamoeba histolytica.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:5118-5122(1989).
CC -1- SIMILARITY: 88% IDENTITY WITH PROTEIN FROM NON-PATHOGENIC
CC STRAIN.
CC Cc
DR PIR; A32935; A32935.
KW Antigen.
FT
SO SEQUENCE 639 AA: 72086 MW: 4752187AB956AEB5 CRC64;

Query Match	9.88;	Score 83.5;	DB 1;	Length 639;
Best Local Similarity	23.68;	Pred. No. 3.1;		
Matches 38;	Conservative 32;	Mismatches 58;	Indels 33;	Gaps 7

```
Qy      1 PVAWNGQNLVLDLSTQIFCHNDYPETITDYVTLORGSAYGVLSNFSGCYAKYSSSI-PR 59
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db     28 PIDISIEDGLN-TTQPIIIND-----TEKISGPGMILRSDPTT-FTNPSFYVTF 76

Qy      60 PTTSETPPRVVYNRSDDKPMPVALLYTLPVSSAGCLYIKAGSLIAVLRLQTNNKNSDDPOF 119
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      77 SNTGRAPLIINYNITTNEEMNSYL----RNAGNAEALERTGNNRYLVLTSRNISREDAOY 131D

Qy     120 VNNIIYANDVVYPFGGCDVASRDYVTLPLDYGSPFIPLY 160
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db    132 ISDFWLK-----AISISNYAVATLEN-----IPITL 156
```

RESULT 10

```

FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 519 AA; 55777 MW; DB4EA526f684740 CRC64;

Query Match
Best Local Similarity 9.8%; Score 83; DB 1; Length 519;
Matches 39; Conservative 19; Mismatches 49; Indels 30; Gaps 8

OY 7 ONLVVLSTQIFCHNXP-----ETITDYVITQKRSATGCVLSNNGTY-----KYSQS 55
DB 249 QPLPLVD-STQIFPAQRISVLANQVGN-WRANPNNGTV--GPAGQINSALILRYGQA 304
OY 56 SYPEPTSETPRVAVNSRTDKPVPALVLYTPVSSAGGLVYKAGSLTAVLILROTNVNSD 115
DB 305 PVAEPPTTQTPSVILPIETNL-HPLARMPVPGSPPTGGVDKA-----LNLAFNNGT 355
OY 116 DFOFVNNTIANDVVP 132
DB 356 NF-----FINNATETP 366

RESULT 11
HEMA_VACCT
ID HEMA_VACCT STANDARD: PRT; 315 AA.
AC P16561.
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Hemagglutinin precursor.
GN HA OR A56R.
OS Vaccinia virus (strain Tian Tan).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10253;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=89328331; PubMed=2754392;
RA Dongyan J., Zhiliang L., Qi J., Hao Y., Yunde H.;
RT "Vaccinia virus hemagglutinin. A novel member of the immunoglobulin
superfamily. ";
RL J. Exp. Med. 170:571-576(1989).
CC
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PPM: THE MAJOR 85 KDA GLYCOPROTEIN IS PHOSPHORYLATED AND SULFATED.
CC O-LINKED GLYCOSYLATION IS REQUIRED FOR HEMAGGLUTINATION AND
CC HEMADSORPTION ACTIVITIES OF INFECTED CELL MEMBRANES.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X15709; CAA33740.1; -.
DR PIR; J10108; HNAVZY.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00409; Ig_1.
KW Hemagglutinin; Envelope protein; Late protein; Glycoprotein; Signal;
KW Transmembrane; Immunoglobulin domain; Phosphorylation; Sulfation.
FT SIGNAL 1 16
FT CHAIN 17 315 HEMAGGLUTININ.
FT DOMAIN 17 279 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 280 303 POTENTIAL.
FT DOMAIN 304 315 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 17 121 IG-LIKE V-TYPE DOMAIN.
FT DISULFID 34 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 337 37 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 315 AA; 34772 MW; E049448640879FF4 CRC64;
Query Match 9.78; Score 82.5; DB 1; Length 315;
Best Local Similarity 23.88; Pred. No. 1.6;
Matches 38; Conservative 22; Mismatches 43; Indels 57; Gaps 8;
OY 4 NVGNLVAVDSTOFCHNDYPTETDYVTLQRGSAVYGSFVYKX----- 52
DB 159 NLNCSVFPEIAT-----PEPITDNV-----EDHDTVTYVSDSINTVSAT 198
OY 53 SGSSYPTTSEPRVYNSRTDKPWPVALYLPVSSAGLVIRAGSLIAVLIRDTNNY 112
DB 199 SGES-----TDERPEPTDKEDHTVTDVSYTSTSSGV-----TTKS 240
OY 113 NSDDFOFVWNIYANDVYVPTG-GCDVSARDVTLDPYR 151
DB 241 TTDAD-LYDTYNDNDVPTVGC-----TTSISNYK 273
RESULT 12
HEMA_IADH2 STANDARD; PRT; 550 AA.
ID HEMA_IADH2
AC P1133; Q84019; Q84020;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1989 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain] (fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/81/78).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11497;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88101364; Pubmed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China."
RT VIROLOGY 162:160-166(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: M19057; AAA43212.1; -
DR PIR: B29971; HMTVS3.
DR HSSP: P03437; 2VIT.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Hemagglutinin; Envelope protein; Glycoprotein.
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 1 350 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61437 MW; 1F2A7E78C31C8 CRC64;
Query Match 9.68; Score 81.5; DB 1; Length 550;
Best Local Similarity 24.38; Pred. No. 3.9;
Matches 36; Conservative 19; Mismatches 58; Indels 35; Gaps 6;
OY 21 NDYPTETDYVTLQRGSAVYGSFVYKX-----NFGSTVYSGSSYPTTSEPRVYNSRT 74
DB 96 NCYPTDVPDVALSLVASSGTFTEGFTWGTONGSN-----ACKRGANGFSRL 151
OY 75 DKPWPVALYLPVSSAGLVIRAGSLIAVLIRDTNNYNSDDFOFVWNIY-----ANN 127
DB 152 N-W-----LTKSGTFVPLNVMTPNNDNSDKL-YINGVHHPSTNOBQT 193
OY 128 DVVPTGCDVSARDVTLDPYRASP 155
DB 194 LYVQASGRVYVSTKRSOOTMIPNAGSRP 221
RESULT 13
HEMA_IADH2 STANDARD; PRT; 550 AA.
ID HEMA_IADH2
AC P12583; Q84011;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain] (fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/8/80).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11358;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; Pubmed=2440178;
RA Kida H., Kawakita Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild ducks."
RT VIROLOGY 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: M16738; AAA3144.1; -
DR PIR: B27813; HMTV80.
DR HSSP: P03437; 2VIT.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 1 350 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61659 MW; A107023ACC9CC353 CRC64;

	Query Match	9.5%	Score 80.5;	DB 1;	Length 550;	
	Best Local Similarity	23.9%;	Pred. No. 4.9;			
	Matches 34; Conservative	24;	Mismatches 61;	Indels 23;	Gaps 5;	
OY	21 NDVEETIDVLTTLRGSAGVGVLNFSGCTVRYSGSSYPFPTSETPRRVYNRRDKDPMPV 80 : : : : : : : : : : : : : : : : : : : NCLPYDVPDYASLRS-----LVASSGTLEKFTIGCFTWGVQTONG---GSKACKRGPA 144					
Dy	81 ALYLTPSSAGGLGYIKAGSLAVLLRGTNNYNSDDPFQVNVI-----ANDVVYPT 133 : 145 SGFFSRRL----WLTKSSTYPVLNVMTPMNDNEFKL-YIVGVHHPSTNOEQLNLVVAS 199					
OY	134 GGCDSARDVTYTLPDYRGSPV 155 : 200 GRATVSTRSOOTIIPNGISRP 221					
	RESULT 14					
	HEMA_VACCV STANDARD; PRT: 314 AA.					
ID	HEMA_VACCV STANDARD; PRT: 314 AA.					
AC	001218:					
DT	01-APR-1993 (Rel. 25, Created)					
DT	01-APR-1993 (Rel. 25, Last sequence update)					
DT	15-JUL-1999 (Rel. 38, Last annotation update)					
DE	Hemagglutinin precursor. HA OR A56R OR SALGR.					
OS	Vaccinia virus (strain WR).					
OC	Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae; Orthopoxvirus.					
NX	NCBI_Taxid=10254;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	MEDLINE=91259063; PubMed=2045793; Smith G.L., Chan Y.S., Howard S.T.; "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near the right inverted terminal repeat." J. Gen. Virol. 72:1349-1376(1991). [2] SEQUENCE FROM N.A. MEDLINE=92410621; Pubmed=1529542; Cavallaro K.F., Esposito J.J.; "Sequences of the raccoon poxvirus hemagglutinin protein."; Virology 190:434-439(1992). RT RL - SUBCELLULAR LOCATION: Type I membrane protein. CC CC - PMW: THE MAJOR 85 kDa GLYCOPROTEIN IS PHOSPHORYLATED AND SULFATED. O-LINKED GLYCOSYLATION IS REQUIRED FOR MEMBRANE ANCHORAGE AND HEMADSORPTION ACTIVITIES OF INFECTED CELL MEMBRANES. CC CC - SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN. CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).					
DR	EMBL; D11079; BAA01829.1; -					
DR	EMBL; M93956; AAA48252.1; -					
DR	PIR; R43381; HNZVZW.					
DR	PIR; Q01793; Q01793.					
DR	InterPro; IPRO03599; IG.					
DR	Interpro; IPRO03006; IG_MHC.					
DR	Pfam; PF00047; Ig_1.					
KM	SMART; SM00409; IG_1.					
RM	Hemagglutinin; Envelope protein; Late protein; Glycoprotein; Signal; Transmembrane; Immunoglobulin domain; Phosphorylation; Sulfation.					
KW	SIGNAL 16					
FT	CHAIN 17 314					
FT	DOMAIN 17 278					
FT	TRANSMEM 279 302					
FT	DOMAIN 303 314					
	CYTOLASMIC (POTENTIAL).					
	POTENTIAL.					
	EXTRACELLULAR (POTENTIAL).					
	POTENTIAL.					
	CYTOLASMIC (POTENTIAL).					

[illegible]

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X87823; CAA61088.1; -
DR EMBL: L27827; AAA17893.1; -
DR HSSP: P02699; 1F88.
DR MGD: MGI:89268; Tncr3.
DR InterPro: IPR000276; GPCR_Rhodopsin.
DR Pfam: PF00001; 7tm_1; 1.
DR PROSITE: PS00237; G-PROTEIN_RECEP_F1.1; 1.
DR PROSITE: PS0262; G-PROTEIN_RECEP_F1.2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KM Phosphorylation; Lipoprotein; Palmitate.
FT DOMAIN 1 71
FT TRANSSEM 72 94
FT DOMAIN 95 104
FT TRANSSEM 105 126
FT TRANSSEM 127 146
FT TRANSSEM 147 168
FT TRANSSEM 169 188
FT TRANSSEM 189 209
FT TRANSSEM 210 232
FT TRANSSEM 233 257
FT TRANSSEM 258 286
FT TRANSSEM 287 308
FT TRANSSEM 309 321
FT TRANSSEM 322 346
FT TRANSSEM 347 385
FT CARBOHYD 9 9
FT CARBOHYD 23 23
FT CARBOHYD 40 40
FT CARBOHYD 60 60
FT DISULFID 145 220
FT LIPID 361 361
FT CONFLICT 198 198
FT CONFLICT 267 267
FT NON_TER 385 385
SQ SEQUENCE 385 AA; 43775 MW; 8A6974E0A455002C CRC64;

Query Match 9.4%; Score 80; DB 1; Length 385;
Best Local Similarity 30.3%; Pred. No. 3.5;
Matches 30; Conservative 15; Mismatches 22; Indels 32; Gaps 6;

QY 27 ITDVTYLRGSAAGVYLSNFGSTVYSSGSPFTTSETPHYVNSRTDK----PWPAVL 82
DB 30 ITTTLALQAG-----NFS-----SALGLPVTQAPSCVDRDLTNOYVPSWRIAL 74

QY 83 YLTPVSSAGLVYKA--GSLTAVLI-----LRQTNV 112
DB 75 W-----SLAYGLVAVVAVFGNLIVILLAHKRMRTVTV 109

RESULT 16
ID PLYD_ASPNG STANDARD: PRT; 373 AA.
AC P22864;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pectin lyase D precursor (EC 4.2.2.10) (PLD) (Pectin lyase I) (PLI).
GN PELD.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutrophiales; Trichocommataceae; mitosporic Trichocommataceae; Aspergillius.
OX NCBI_TaxID=5061;
RN NCBI_SEQUENCE FROM N.A., AND SEQUENCE OF 20-30 AND 148-167.
RP STRAIN=N756;
RX MEDLINE=90323592; PubMed=2373363;
RA Gysler C., Hansen J.A.M., Kester H.C.M., Visser J., Helm J.;
RT "Isolation and structure of the pectin lyase D-encoding gene from
RT Aspergillus niger.";

RL Gene 89:101-108(1990).
CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectin to give
CC oligosaccharides with terminal 4-deoxy-6-methyl-alpha-D-galact-4-
CC enuronosyl groups.
CC -1- PTM: N-ACETYLDUCOSAMINE WAS NOT FOUND IN PURIFIED PECTIN LYASE,
CC BUT SOME O-GLYCOSYLATION MAY OCCUR.
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M55657; AAA2701.1; -
DR PIR: JH0155; JH0155.
DR HSSP: O01172; 1IDJ.
DR InterPro: IPR002022; Amb_allergen.
DR Pfam: PF00544; pec_lyase; 1.
KW Lyase; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 19
FT CHAIN 20 373
FT ACT_SITE 255 255
FT CARBOHYD 128 128
FT CARBOHYD 274 274
FT CARBOHYD 348 348
FT CONFLICT 27 27
SQ SEQUENCE 373 AA; 39038 MW; B758DCCFEF2345C CRC64;

Query Match 9.3%; Score 79; DB 1; Length 373;
Best Local Similarity 23.6%; Pred. No. 4.2;
Matches 38; Conservative 25; Mismatches 72; Indels 26; Gaps 7;

QY 9 LVVDLSQIFCHNDYPTTIDYVTLRGSAAGVYLSNFGSTVYSSGSPFTTSETPR 67
DB 200 LGTDADRSVSTNNYNGESDYSATCDGHYNNVYLDGSSDQKVFSG-NLYKTSGRAPK 258

QY 68 V-----VYNSRTDKPMPVALYLPVSSAGLVYKASILA-VLIRQTN-----YNS 114
DB 259 VQDNTVHTIYNNVNNSGAFET-----GSGGYLABGNFNSVDYLEIDTFEGALFSS 314

QY 115 DDQFVNNIYANNDDVYVPTGCGDVASRDVY-----TLP 148
DB 315 DSASTCESYIGRSCVANVNGDLTGSTVYLSNLGDTLP 355

RESULT 17
ID HEMA_IADH5 STANDARD: PRT; 550 AA.
AC P12586; O84015; O84016;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/21/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11361;
RN NCBI_SEQUENCE FROM N.A.
RP MEDLINE=87265458; PubMed=2440178;
RX Kida H., Kawoka Y., Naeve C.W., Webster R.G.;
RA "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -! SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16741; AAA43147.1; -
DR PIR: E27813; HMTV21.
DR HSSP: P03437; 2V1U.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR Prodom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 7 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 7 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 8 7 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 179 179 YV -> VI (IN PIR DATA BANK).
FT CONFLICT 388 388 K -> T (IN PIR DATA BANK).
SQ SEQUENCE 550 AA; 61856 MW; 48401C867A15BFC CRC64;

Query Match 9.3%; Score 78.5; DB 1; Length 550;
Best Local Similarity 25.0%; Pred. No. 7.5;
Matches 37; Conservative 18; Mismatches 58; Indels 35; Gaps 7;

QY 21 NDYPTITDYTLQKSGAYGVLS-----NFSGTIVYSGSYFPPTTSEPRVYNSRT 74
DB 96 NCYPTVDVYASLNLVASSGTLEFTEGFTWTGTONGSN---ACKRGPSGFFSRL 151
DB 75 DKPVPALYLPVSSAGLVTKAGSLIYLLRQTNVNSDDFOFVNIT-----ANN 128
QY 152 N-W-----LTKSGSTYVLYNTPMNDNPKL-YVGVHHPSINQEDTN 193
DB 129 VVPTGCGDYASADVTLTPDYGSPV 155
DB 194 LYVQASGRVTVSTRSQOTIIPNIGSRP 221

RESULT 18
HEMA_IADH7 STANDARD; PRT; 550 AA.
ID HEMA_IADH7 STANDARD; PRT; 550 AA.
AC P12588; Q84018; Q89470;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/10/85).
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzaviridae;
OC NCBL_TaxID=11365;
OX NCBL_TaxID=11365;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawakita Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks".
RL VIROLOGY 159:109-119(1987).
CC -! FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -! SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -! SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16743; AAA43149.1; -
DR PIR: G27813; HMTV15.
DR HSSP: P03437; 3HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR Prodom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 7 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 7 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 8 7 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61761 MW; 6EF81793281D53B8 CRC64;

Query Match 9.3%; Score 78.5; DB 1; Length 550;
Best Local Similarity 24.3%; Pred. No. 7.5;
Matches 36; Conservative 18; Mismatches 59; Indels 35; Gaps 6;

QY 21 NDYPTITDYTLQKSGAYGVLS-----NFSGTIVYSGSYFPPTTSEPRVYNSRT 74
DB 96 NCYPTVDVYASLNLVASSGTLEFTEGFTWTGTONGSN---ACKRGPSGFFSRL 151
DB 75 DKPVPALYLPVSSAGLVTKAGSLIYLLRQTNVNSDDFOFVNIT-----ANN 127
QY 152 N-W-----LTKSGSTYVLYNTPMNDNPKL-YVGVHHPSINQEDTN 193
DB 128 DVVPTGCGDYASADVTLTPDYGSPV 155
DB 194 LYVQASGRVTVSTRSQOTIIPNIGSRP 221

RESULT 19
HEMA_IAX31 STANDARD; PRT; 249 AA.
ID HEMA_IAX31 STANDARD; PRT; 249 AA.
AC P03438;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain X-31).
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzaviridae;
OC NCBL_TaxID=11469;
OX NCBL_TaxID=11469;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81030852; PubMed=7421990;
RA Getling M.-J., Bye J., Skehel J.J., Waterfield M.;
RT "Cloning and DNA sequence of double-stranded copies of haemagglutinin
RT genes from H2 and H3 strains elucidates antigenic shift and drift in
RT human influenza virus".
RL Nature 287:301-306(1980).
CC -! FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -! SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -! SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR PIR: A04051; HMTVH.

```

DR PIR: A93233; A93233.
DR HSSP: P03437; 1HGJ.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00309; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
DR Envelope protein; Hemagglutinin; Glycoprotein; signal.
FT SIGNAL 1 16
FT CHAIN 17 >249
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 249 249
SQ SEQUENCE 249 AA: 27373 MM: 9537AA970BB79183 CRC64;

Query Match 9.1%; Score 77.5; DB 1; Length 249;
Best Local Similarity 25.7%; Pred. No. 3.6;
Matches 44; Conservative 21; Mismatches 67; Indels 39; Gaps 9;

QY 1 PVVAVGQNLVVDL---STQIFCHNDYPETITDYTLQRGSAVGVLS-----NFGSIVK 51
DB 90 PHCDVFQNETWDLFVRSKAF-SNCCYPDVPDASLRSVLASSGTLFTTEGFTWGTQ 148
QY 52 YSGSSYPEPTTSETPRVVYNSRTDKPMPVALYLPVSSAGGLVAKGLIYVLIIRQTNN 111
DB 149 NGGSN----ACKRGPSSGFSRLN--W-----LTKSGSTYVVLVMTMPNN 187
QY 112 YNSDFOFVNNIY---ANND-----VVVPTGGCDVSARDVTLTLPDYGSPV 155
DB 188 DNFKL-YIMGIIHPRSTNOEQTSLYVQSGRVYSTRSQGTIIIPNIGSRP 237

RESULT 20
HEMA_VACC
ID HEMA_VACC STANDARD; PRT; 315 AA.
AC P20978;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor.
GN HA OR A56R.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
NC NCBI_Taxid=10249;
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkins M.E., Winslow J.P.,
RT "The complete DNA sequence of vaccinia virus.",
RL Virology 179:247-266(1990).
RN COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
RT "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL Virology 179:517-563(1990).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PM: THE MAJOR 85 kDa GLYCOPROTEIN IS PHOSPHORYLATED AND SULFATED.
CC O-LINKED GLYCOSYLATION IS REQUIRED FOR HEMAGGLUTINATION AND
CC HEMADSORPTION ACTIVITIES OF INFECTED CELL MEMBRANES.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: M35027; AAA48191.1; -.
DR PIR: D42523; HNVZ4X.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003066; Ig_MHC.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00409; Ig; 1.
DR SMART: SM00409; Ig; 1.
DR Hemagglutinin; Envelope protein; Late protein; glycoprotein; signal;
KW Transmembrane; Immunoglobulin domain; Phosphorylation; Sulfation.
FT SIGNAL 1 16
FT CHAIN 17 315
FT DOMAIN 17 279 HEMAGGLUTININ (BY SIMILARITY).
FT TRANSMEM 280 303 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 304 315 POTENTIAL.
FT DOMAIN 17 121 CYTOPLASMIC (POTENTIAL).
FT DISULFID 34 103 IG-LIKE V-TYPE DOMAIN.
FT CARBOHYD 37 37 POTENTIAL.
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 315 AA: 34778 MM: C7EED5EA12134D3 CRC64;

Query Match 9.1%; Score 77.5; DB 1; Length 315;
Best Local Similarity 21.3%; Pred. No. 4.8;
Matches 39; Conservative 22; Mismatches 53; Indels 69; Gaps 8;

QY 2 VVAVGQNLVVDLSTQIFCHN-----DY-----PETITDYTLQRGSAV 39
DB 127 IVNIDSESTIDILSGTSHSPETSEKPDYIDNSCSSVFETAPETIDV----- 178
QY 40 GGVLSNFGVYK-----SGSSYPEPTTSETPRVVYNSRTDKPMPVALYLPVS 88
DB 179 ----EDHTDYITSDSINTVSATGES----TDEPTEPTIDKEEDHTVDTYSYTVS 230
QY 89 SAGGLVTKAGSLIYVLIIRQTNNNSDFOFVNNIYANDVVPVPGCDVSARDVYVTLP 148
DB 231 TSSGIV-----TTKSTTDAD-LYDYNNDVTPSY-----TVGSSYTSIS 270
QY 149 DYR 151
DB 271 NYK 273

RESULT 21
HEMA_IADH1
ID HEMA_IADH1 STANDARD; PRT; 550 AA.
AC P12582; Q84021; Q84022;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/5/77).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
NC NCBI_Taxid=11357;
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawoka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M16737; AAA43143.1; .
DR PIR: A27813; HMTV77.
DR HSSP: P03437; 3HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutin; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 330 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61705 MW; 7E7ACFE716FC969A CRC64;

Query Match 9.1%; Score 77.5; DB 1; Length 550;
Best Local Similarity 24.3%; Pred. No. 9.3;
Matches 36; Conservative 18; Mismatches 59; Indels 35; Gaps 6;

OY 21 NDPEPTIDYVTLQGSAYGVLS-----NESTGVKSGSSYPPTTSEPRVYNSRT 74
DB 96 NCYPDVPDYASLRSLVASSGLEITGEFTWTGYTONGSN-----ACKRGASGFPSRL 151
OY 75 DKPWPVALYLPVSSAGGLVTKAGSLVALLRQTNNTNSDFOFWMNY-----ANN 127
DB 152 N-W-----LTKSGSTPYLVNTYMPNNDNPKL-YIMGVHHPTDQEQTN 193
OY 128 DVVPTGGCDVSARDVYTLDPYRGSSVP 155
DB 194 LYVQASGRVTVSTRSQGTIIIPNIGSRP 221

RESULT 22
HEMA_IADH4 STANDARD; PRT; 550 AA.
AC P12585; Q84013; Q84014;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/7/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11360;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87265458; PubMed=2440178;
RX Kida H., Kawoka Y., Neeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks";
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL: M16740; AAA43146.1; .
DR PIR: D27813; HMTV89.
DR HSSP: P03437; 2VTV.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutin; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 330 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61664 MW; A16B2CFC8CBBBD9D0 CRC64;

Query Match 9.1%; Score 77.5; DB 1; Length 550;
Best Local Similarity 24.3%; Pred. No. 9.3;
Matches 36; Conservative 18; Mismatches 59; Indels 35; Gaps 6;

OY 21 NDPEPTIDYVTLQGSAYGVLS-----NESTGVKSGSSYPPTTSEPRVYNSRT 74
DB 96 NCYPDVPDYASLRSLVASSGLEITGEFTWTGYTONGSN-----ACKRGASGFPSRL 151
OY 75 DKPWPVALYLPVSSAGGLVTKAGSLVALLRQTNNTNSDFOFWMNY-----ANN 127
DB 152 N-W-----LTKSGSTPYLVNTYMPNNDNPKL-YIMGVHHPTDQEQTN 193
OY 128 DVVPTGGCDVSARDVYTLDPYRGSSVP 155
DB 194 LYVQASGRVTVSTRSQGTIIIPNIGSRP 221

RESULT 23
HEMA_IADHK STANDARD; PRT; 550 AA.
AC P43257;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hong Kong/7/75).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11364;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=91341491; PubMed=1875195;
RX Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
RT of avian H3 influenza viruses to pigs in southern China, where the
RT A/hong kong/68 (H3N2) strain emerged.";
RL J. Gen. Virol. 72:2007-2010(1991).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D00929; BAA00769.1; -
DR HSSP: P03437; 2V1U.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61549 MW; 8646398829FELBA9 CRC64;
Query Match 9.1%; Score 77.5; DB 1; Length 550;
Best Local Similarity 24.7%; Pred. No. 9.3;
Matches 36; Conservative 17; Mismatches 58; Indels 35; Gaps 6;
QY 23 YPETIDYVTLQKSAVGVL-----NFGTVKYSGSSTPEPTSEPRVYNSRTDK 76
DB 98 YPYDVPYASLRSLVASSGLEFTTEGFTWGTQNGSS-----ACKRGASGFFSRLN- 152
QY 77 PWPVALYLPVSSAGLVIRAGSLIAVLIRQTNNVNSDDQFVWNLY-----ANNV 129
DB 153 -W-----LTRSGSTYPVLNVTMPNNDKIL-YIMGVHHPSTNOEQTNLY 195
QY 130 VVPTGCDVSARDVYVTLPPYRGVSP 155
DB 196 VQASGRVTVSTRSQOTIIPNIGSRP 221
RESULT 24
HEMA_IJAZH3 STANDARD; PRT; 550 AA.
ID HEMA_IJAZH3
AC P11134; Q84025; Q84026;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/12/6/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11498;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88101364; Pubmed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China."
RL Virology 162:160-166(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M19056; AAA43211.1; ALT_TERM.
DR PIR: A29971; HMTVS2.
DR HSSP: P03437; 2V1U.
DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Hemagglutinin; Envelope protein; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61580 MW; 991P6DB8C02F24F2 CRC64;
Query Match 9.1%; Score 77.5; DB 1; Length 550;
Best Local Similarity 24.7%; Pred. No. 9.3;
Matches 36; Conservative 17; Mismatches 58; Indels 35; Gaps 6;
QY 23 YPETIDYVTLQKSAVGVL-----NFGTVKYSGSSTPEPTSEPRVYNSRTDK 76
DB 98 YPYDVPYASLRSLVASSGLEFTTEGFTWGTQNGSS-----ACKRGASGFFSRLN- 152
QY 77 PWPVALYLPVSSAGLVIRAGSLIAVLIRQTNNVNSDDQFVWNLY-----ANNV 129
DB 153 -W-----LTRSGSTYPVLNVTMPNNDKIL-YIMGVHHPSTNOEQTNLY 195
QY 130 VVPTGCDVSARDVYVTLPPYRGVSP 155
DB 196 VQASGRVTVSTRSQOTIIPNIGSRP 221
RESULT 25
HEMA_IJAZH3 STANDARD; PRT; 566 AA.
ID HEMA_IJAZH3
AC P03437;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Aichi/2/68).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=150147;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80254693; Pubmed=7402351;
RA Verhoeven M., Fang R., Min Jou W., Devos R., Huylebroeck D., Saman E., Fiers W.;
RT "Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/Aichi/2/68 and A/Victoria/3/75."
RL Nature 286:771-776(1980).
RN [2]
RN X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE=81123029; Pubmed=7464906;
RA Wilson I.A., Skehel J.J., Wiley D.C.;
RT "Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3-A resolution."
RL Nature 289:366-373(1981).
RN [3]
RN X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=88232903; Pubmed=3374584;
RA Weis W.I., Brown J.H., Cusack S.C., Paulson J.C., Skehel J.J., Wiley D.C.;
RT "Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid."
RL Nature 333:426-431(1988).
RN [4]
RN X-RAY CRYSTALLOGRAPHY OF A MUTANT WITH GLY-457.
RX MEDLINE=90107940; Pubmed=2295311;
RA Weis W.I., Cusack S.C., Brown J.H., Daniels R.S., Skehel J.J., Wiley D.C.;

FT HELIX 505 515
SQ SEQUENCE 566 AA; 63415 MW; E395659C23CAFECA CRC64;

Query Match 9.1%; Score 77.5; DB 1; Length 566;
Best Local Similarity 25.7%; Pred. No. 9.6;
Matches 44; Conservative 21; Mismatches 67; Indels 39; Gaps 9;

QY 1 PNVNQGQNLVVDL---STQFCHNDPFTITDVTQROGSAYGVLS-----NFGSTVK 51
DB 90 PHCDVQNETWDLFVERSKAF-SNCYPYDVPDASLRSLVASSGTLFEFTGFTWGVQ 148
QY 52 YSSGSPFTTSETPRVYVSRDTPPVVALYLPVSSAGGLVIKAGSLIAVLILROTN 111
DB 149 NGGSN-----ACKRGPGSGFFSRLLN--W-----LTKSGSTYPVLNTPMNP 187
QY 112 YNSDDEQFVWNIY---ANND---VVVPTGGCDVSADVTVLPDYRGVSP 155
DB 188 DNFDKL-YIWGHHPTNQSTSLYQASGRVTVSTRSQOTIIPNIGSRP 237

RESULT 26
HEMA_IADHL STANDARD; PRT; 550 AA.

ID HEMA_IADHL

AC P43258;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;

HA. Hemagglutinin HA2 chain] (Fragment).

OS Influenza A virus (strain A/Duck/Hong Kong/64/76).

OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza A viruses; Influenzavirus A.

NCBI_TaxID=45412;

RNA [1]

SEQUENCE FROM N.A.

RP MEDLINE=91341491; Pubmed=1875195;

RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;

"Molecular evidence for a role of domestic ducks in the introduction

of avian H3 influenza viruses to pigs in southern China, where the

A/Hong Kong/68 (H3N2) strain emerged."

J. Gen. Virol. 72:2007-2010(1991).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

(HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: D00937; BAA00771.1; -

DR HSSP: P03437; 2YIU.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR ProDom: PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein.

FT NON_TER 1 1

FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.

FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.

FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 550 AA; 61718 MW; A51C56789E4BE9A CRC64;

Query Match 9.0%; Score 76.5; DB 1; Length 550;
Best Local Similarity 24.3%; Pred. No. 11;
Matches 36; Conservative 18; Mismatches 59; Indels 35; Gaps 6;

QY 21 NDYPTITDVTYLQSGAYGVLS-----NFGSTVKYSGSSPTTETPRVYNSRT 74
DB 96 NCPYDVPDYPASLRSVLVASSGTLFEFTGFTWGVQNGGSN-----ACKRGPGNGFFSRLL 151
QY 75 DKRWPAALTLTPSSAGGLVIKAGSLIAVLILROTNYSDDQFQWNIY-----ANN 127
DB 152 N-W-----LTKSGSTYPVLNTPMNPNDKLL-YIWGHHPTNQSTSLYQASGRVTVSTRSQOTIIPNIGSRP 193
QY 128 DVVPTPGCDVSADVTVLPDYRGVSP 155
DB 194 LVYQASGRVTVSTRSQOTIIPNIGSRP 221

RESULT 27
HEMA_IAGHK STANDARD; PRT; 550 AA.

ID HEMA_IAGHK

AC P43260;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;

HA. Hemagglutinin HA2 chain] (Fragment).

OS Influenza A virus (strain A/Goose/Hong Kong/10/76).

OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza A viruses; Influenzavirus A.

NCBI_TaxID=45414;

RNA [1]

SEQUENCE FROM N.A.

RP MEDLINE=91341491; Pubmed=1875195;

RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;

"Molecular evidence for a role of domestic ducks in the introduction

of avian H3 influenza viruses to pigs in southern China, where the

A/Hong Kong/68 (H3N2) strain emerged."

J. Gen. Virol. 72:2007-2010(1991).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

(HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: D00930; BAA00770.1; -

DR HSSP: P03437; 2YIU.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR ProDom: PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein.

FT NON_TER 1 1

FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.

FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.

FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 550 AA; 61676 MW; 9A1E094DA28BACD2 CRC64;

Query Match 9.0%; Score 76.5; DB 1; Length 550;
Best Local Similarity 24.3%; Pred. No. 11;
Matches 36; Conservative 18; Mismatches 59; Indels 35; Gaps 6;

[illegible]


```

-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
CC  EMBL: 249811; CAA89962.1; -
CC  KW Hypothetical protein; Transmembrane.
CC  FT TRANSMEM 17 33 POTENTIAL.
CC  SQ SEQUENCE 371 AA; 43688 MW; 890BAD4A0E66379B CRC64;
CC
Query Match          9.0%; Score 76; DB 1; Length 371;
Best Local Similarity 21.3%; Pred. No. 8;
Matches 33; Conservative 27; Mismatches 51; Indels 44; Gaps 6;
CC
OY  9 LVLDSTQICNDYPTIDYVTLQSGAY---GGVLNFGSTVYKSGSSYPPTSE 64
DB  24 LIYMTLVFNGHGYKOTLDRLSLKNNPEENDNAVLEKPEPKYTM-SLFTMPSTEE 82
OY  65 -----TPRVVNSRTDKPWPVALYLTVPSSAGGLVYKAGSLAVILKOTNNYNS 114
DB  83 DYFNATRVLIHRLKYHPTTKSKYP-----IHLALRGVDEMKI 121
OY  115 DDFQ-----FVNNITVANDVYVPTGCG--DVSNR 141
DB  122 ERFKRDGASVYIVDPIASSDIVYDTSSFSQSIAR 156
CC
RESULT 30
ILVD_SULSO          STANDARD;          PRT;          558 AA.
ID  ILVD_SULSO
AC  Q970B2;
DT  15-JUN-2002 (Rel. 41, Created)
DT  15-JUN-2002 (Rel. 41, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Dihydroxy-acid dehydratase (EC 4.2.1.9) (DAD).
GN  ILVD OR SS03107.
OS  Sulfolobus solfataricus.
OC  Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC  Sulfolobus.
OX  NCBI_TaxID=2287;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 35092 / DSM 1617 / P2;
RX  MEDLINE=21332296; PubMed=11427726;
RA  She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA  Aweyer M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,
RA  De Moers A., Ernuso G., Fletcher C., Gordon P.M.K.,
RA  Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA  Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA  Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA  Garrett R.A., Regan M.A., Sensen C.W., Van der Oost J.;
RT  "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RT  Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
RL  [1]
CC  -1- CATALYTIC ACTIVITY: 2,3-dihydroxy-3-methylbutanoate = 3-methyl-2-
CC  oxobutanoate + H(2)O.
CC  -1- COFACTOR: Binds 1 4Fe-4S cluster (potential).
CC  -1- PATHWAY: Valine and isoleucine biosynthesis; fourth step.
CC  -1- SIMILARITY: BELONGS TO THE ILVD / EDD FAMILY.
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
CC  EMBL: AE006901; AAK43209.1; -
CC  InterPro: IPR004404; ILVD.

```

```

DR  InterPro: IPR000581; ILVD_EDD_family.
DR  Pfam: PF00920; ILVD_EDD_1.
DR  ProDom: PD002691; ILVD_EDD_family; 1.
DR  TIGRfam: TIGR00110; ILVD_1.
DR  PROSITE: PS00886; ILVD_EDD_1; 1.
DR  PROSITE: PS00887; ILVD_EDD_2; 1.
KW  Branched-chain amino acid biosynthesis; lyase; Iron-sulfur; 4Fe-4S;
KW  Complete proteome.
FT  METAL 123 123 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT  METAL 195 195 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ  SEQUENCE 558 AA; 59421 MW; 059B2D16BAC62043 CRC64;
CC
Query Match          9.0%; Score 76; DB 1; Length 558;
Best Local Similarity 29.7%; Pred. No. 13;
Matches 30; Conservative 10; Mismatches 33; Indels 28; Gaps 4;
CC
OY  33 LQRCGAVGVLSNFGSTVYKSGSSYPPTT--SETPRVVNSRTDKPWPVALYTPVSSA 90
DB  332 LDAGLLHGDVLTVGKTKMKNLEQYKYPNPHSHIVDVKN-----PIKPR 377
OY  91 GGAVIKAGSLI---AVLILKQTN-----NYSNDDQFQ 119
DB  378 GGAVILKGLAPGAVIKVATINAVKFEKGAKAYNSDDAF 418
CC
RESULT 31
PMIL_CHLPN          STANDARD;          PRT;          928 AA.
ID  PMIL_CHLPN
AC  O86164; Q9K299;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Probable outer membrane protein pmil1 precursor (Polymorphic membrane
DE  protein 11) (Outer membrane protein 4).
GN  PM11 OR OMP4 OR CPN0449 OR CP0302.
OS  Chlamydia pneumoniae (Chlamydia pneumoniae).
OC  Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX  NCBI_TaxID=83558;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CWL029/VR-1310;
RX  MEDLINE=99081766; PubMed=9864239;
RA  Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
RT  "Identification of two novel genes encoding 97- to 99-kilodalton outer
RT  membrane proteins of Chlamydia pneumoniae.";
RT  Infect. Immun. 67:375-383(1999).
RL  [1]
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=VR1310;
RX  MEDLINE=20007584; PubMed=10539856;
RA  Christiansen G., Boesen T., Hjerno K., Daugaard L., Mygind P.,
RA  Madsen A.S., Knudsen K., Falk E., Birkelund S.;
RT  "Molecular biology of Chlamydia pneumoniae surface proteins and their
RT  role in immunopathogenicity.";
RT  Am. Heart J. 138:S491-S495(1999).
RL  [3]
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CWL029;
RX  MEDLINE=99206606; PubMed=10192388;
RA  Kaiman S., Mitchell W., Karathe R., Lammel C., Fan J., Hyman R.W.,
RA  Ollinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT  "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RT  Nat. Genet. 21:385-389(1999).
RL  [4]
RN  [4]
RP  SEQUENCE FROM N.A.
RC  STRAIN=AR39;
RX  MEDLINE=20150255; PubMed=10684935;
RA  Reed T.D., Brunham R.C., Shen C., Gail S.R., Heidelberg J.F.,
RA  White O., Hickey E.K., Peterson J., Ueberback T., Berry K., Baas S.,
RA  Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA  Givon M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA  Eisen J., Fraser C.M.;
RT  "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia

```

```

RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-0138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shitai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishi K., Hattori M., Kohara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae 0138
RT from Japan and CHL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -i- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -i- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AJ001311; CA04672.1; -
DR EMBL: AJ133034; CAB37072.1; -
DR EMBL: AE001628; AAD18593.1; -
DR EMBL: AE002192; AAF38159.1; ALT_INIT.
DR EMBL: AP002546; BAA98658.1; -
DR PHC1-2DPAGE; 086164; -
DR TIGR: CP0302; -
DR InterPro: IPR003368; Chlamydia_PMP.
DR InterPro: IPR003357; OMP.
DR Pfam: PF02385; OMP. 1.
DR Pfam: PF02415; DUF145; 2.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 1 28 PROBABLE OUTER MEMBRANE PROTEIN PMP11.
SQ SEQUENCE 928 AA; 98903 MW; 788BCDD62C911402 CRC64;

Query Match 9.0%; Score 76; DB 1; Length 928;
Best Local Similarity 23.8%; Pred. No. 24;
Matches 29; Conservative 25; Mismatches 34; Indels 34; Gaps 6;

QY 3 VNVGON----LVVDLSQIFCHANDPE--TITDYVTLQSGASVGVLSNFGVTKYSGS 55
DB 353 INIGSGKFTLEIRAKKNHIFPYDITSEGTSSDVLKINNSA--GALNPGQTLFSGE 410
QY 56 SYPFETETPRVYVNSKDKPMVALYLTIPVSSAGGLVI-----KAGSLI 101
DB 411 T---LTADKLKVDNLKSS-----FTQPVSLSGKLLQKGVLTSTSPSQEAGSIL 459
QY 102 AV 103
DB 460 GM 461

RESULT 32
HEMA_IADA3
ID HEMA_IADA3 STANDARD; PRT; 566 AA.
AC P26134;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
DE Hemagglutinin HAI chain].
GN HA.
OS Influenza A virus (strain A/Duck/Alberta/78/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11348;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114135; PubMed=1731092;

```

```

RA Bean W.J., Schell M., Katz J., Kawoaka Y., Naeve C., Gorman O.,
RA Webster R.G.;
RT "Evolution of the H3 influenza virus hemagglutinin from human and
RT nonhuman hosts.";
RL J. Virol. 66:1129-1138(1992).
CC -i- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -i- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -i- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M73771; -; NOT_ANNOTATED_CDS.
DR HSSP; P03437; 2YIU.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 1 17
FT CDS 17 344
FT CARBOHYD 346 566
FT FT 23 23 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63534 MW; FE19AB6FE9415B89 CRC64;

Query Match 8.9%; Score 75.5; DB 1; Length 566;
Best Local Similarity 25.0%; Pred. No. 15;
Matches 37; Conservative 17; Mismatches 59; Indels 35; Gaps 6;

QY 21 NDYPETIDYVYTLQSGASVGVLS-----NFGTVKYSGSSEFPETSETPRVVNSRT 74
DB 112 NCYPDVADVDSLSNIVASSGTLEIFEGFTWGTQWGSN-----ACKRRPAGFSRL 167
QY 75 DKPWPVALYLPVSSAGGLVTKAGSLAVLTLROTNNYNSDFQVWNIT-----ANN 127
DB 168 N-W-----LTKSGSTYFVLNVTPMNDNFDKL-YVGVHHPSTNOEOTN 209
QY 128 DVVYPTGGCDYSARDVYVTLPDYNGSVP 155
DB 210 LYVQASGRVYSTRSQOTILPNIISR 237

RESULT 33
HEMA_IAMAO
ID HEMA_IAMAO STANDARD; PRT; 566 AA.
AC P26138;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Mallard/New York/68/74/78).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11436;
RN [1]
RP SEQUENCE FROM N.A.
RX Bean W.J., Schell M., Katz J., Kawoaka Y., Naeve C., Gorman O.,

```

RA Webster R.G.:
RT "evolution of the H3 influenza virus hemagglutinin from human and
nonhuman hosts".
RL J. VIROL. 66:1129-1138(1992).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: M73776; -; NOT_ANNOTATED_CDS.
DR HSSP: P03437; 2VIU.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; Hemagglutn12.
DR ProDom: PD000225; Hemagglutn.1.
KW Envelope protein; Hemagglutinin; Glycoprotein; signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63574 MW; BB206011CDBD9A3B CRC64;

Query Match 8.9%; Score 75.5; DB 1; Length 566;
Best Local Similarity 25.0%; Pred. No. 15;
Matches 37; Conservative 16; Mismatches 60; Indels 35; Gaps 6;

QY 21 NDYPTTDTVYTLQSGAYGCVLS-----NFSCTVYKSSSYFFPTTSEPRVYVYVSR 74
DB 112 NCYPYDVPDYASLSLVAASCTLEFTEAFWTGVTONGSG---ACKRPGNGFESRL 167
QY 75 DKHPVALYLPVSSAGGVIRKAGSLAVLILRQTNVNSDDPQFVNY-----ANN 127
DB 168 N-W-----LTKSGSAIPVLTWYKPNNDNFDKL-YVGVHHPSTNOEQTN 209
QY 128 DVVPTGCDVSARDVTTLPLDYRGVVP 155
DB 210 LVYQASGRVYVSTRRSQGTIIIPNGSRP 237

RESULT 34
LAC2_TRAVE STANDARD; PRT; 519 AA.
AC 012718: 012716;
DC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Laccase 2 precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)
DE (uridylol oxidase) (Diphenol oxidase) (Laccase I).
GN LCC2 OR LCCI.
OS Trametes versicolor (White-rot fungus).
OC Eukaryote; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Aphyllophorales; Trametes.
OX NCBI_TaxID=5325;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=521;
RX MEDLINE=97464057; PubMed=9322748;
RA Ong E., Pollock W.B., Smith M.;
RT Cloning and sequence analysis of two laccase complementary DNAs from

RT the ligninolytic basidiomycete Trametes versicolor";
RL Gene 196:113-119(1997).
CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
PRODUCTS (PROBABLE).
CC -1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) -> 4 benzosemiquinone + 2
H(2)O.
CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: U44851; AA86659.1; -;
DR EMBL: U44430; AAC9828.1; -;
DR HSSP: P37064; 1A02.
DR InterPro: IPR001117; Cu-oxidase.
DR Pfam: PF00394; Cu-oxidase.3.
DR InterPro: IPR002355; MulticCu-oxidase2.
DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; FALSE_NEG.
KW Oxidoreductase; signal; Copper; Metal-binding; glycoprotein; Repeat;
KW Lignin degradation; Multigene family.
FT SIGNAL 1 20
FT CHAIN 21 519
FT DOMAIN 22 147 LACCASE 2.
FT DOMAIN 159 301 PLASTOCYANIN-LIKE 1.
FT DOMAIN 368 490 PLASTOCYANIN-LIKE 2.
FT METAL 84 84 PLASTOCYANIN-LIKE 3.
FT METAL 86 86 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 129 129 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 131 131 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 131 131 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 415 415 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 418 415 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 420 420 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 422 422 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 472 472 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 473 473 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 474 474 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 478 478 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 478 478 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 74 74 COPPER (TYPE 1) (BY SIMILARITY).
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 237 237 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 69 69 D -> V (IN REF. 1; AAC9828).
SQ SEQUENCE 519 AA; 55810 MW; BFFB5B4CD0007702 CRC64;

Query Match 8.8%; Score 75; DB 1; Length 519;
Best Local Similarity 27.7%; Pred. No. 15;
Matches 38; Conservative 19; Mismatches 50; Indels 30; Gaps 8;

QY 7 QNLVVDSTQFFCHNDP-----ETITDYVTLQSGAYGCVLSNFSGTV-----KTSGS 55
DB 249 QPLLVLD-SIQIFAQRYSFVIANQTYGN-YVWRANPNFETV--GGAGGISALIRIQGA 304
QY 56 SYPEPTTSEPRVYVYSTRDKPMFVALYLPVSSAGGVIRKAGSLAVLILRQTNVNSD 115
DB 305 PVAEPTTQTQTSVPLLETNL-HPLARMPVPGSPTPGVDNA-----LNLAFNFGCT 355
QY 116 DFGFVWNIYANDVYVP 132

Db 356 NF-----FINNATFTP 366

RESULT 35

PELC_ERWCH STANDARD; PRT; 375 AA.

ID PELC_ERWCH

AC P11073;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Pectate lyase C precursor (EC 4.2.2.2).

GN PELC.

OS Erwinia chrysanthemi.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

CC Pectobacterium.

OX NCBI_TaxID=556;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-EC16;

RX MEDLINE=88298652; PubMed=3042750;

RA Tamaki S.J., Gold S., Robeson M., Manulis S., Keen N.T.;

RT "Structure and organization of the pel genes from Erwinia chrysanthemi EC16.";

RL Bacteriol. 170:3468-3478(1988).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).

RX MEDLINE=93276270; PubMed=8502994;

RA Yoder M.D., Keen N.T., Jurnak F.;

RT "New domain motif: the structure of pectate lyase C, a secreted plant virulence factor.";

RL Science 260:1503-1507(1993).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).

RA Lietzke S.E., Scavetta R.D., Yoder M.D., Jurnak F.A.;

RT "The refined three-dimensional structure of pectate lyase E from Erwinia chrysanthemi at 2.2-A resolution.";

RL Plant Physiol. 111:73-92(1996).

CC -1- FUNCTION: INVOLVED IN MACERATION AND SOFT-ROTTING OF PLANT TISSUE.

CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectate to give oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at their non-reducing ends.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1. PLADES SUBFAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL: M19411; AAA24849.1; -.

DR PIR: A31091; W2WC6C.

DR PDB: 2PEC; 14-FEB-95.

DR PDB: 1AIR; 16-JUN-97.

DR InterPro: IPR002022; Amb_allergen.

DR Pfam: PF00544; pec_lyase; 1.

DR Lyase; Multigene family; Signal; 3D-structure.

FT SIGNAL 1 22

FT CHAIN 23 375 PECTATE LYASE C.

FT DISULFID 94 177

FT DISULFID 351 374

FT ACT_SITE 240 240 POTENTIAL.

SO SEQUENCE 375 AA; 39943 MW; F76DD8195A35B886 CRC64;

Query Match 8.8%; Score 74.5; DB 1; Length 375;

Best Local Similarity 22.4%; Pred. No. 11;

Matches 37; Conservative 12; Mismatches 56; Indels 49; Gaps 8;

QY 12 DLSTQIFCHNDYPTITDYVTLGRGS--AYGVLNFSGRT--VKYSGSSYPTTSET 65

| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 222 DTGRNITYHHNYNDVNARLPLQRGGLVHAVNNLYNTITSGSLWRQGA----- 272

QY 66 PRVYNSRTDKPWPVALYLPVSSA-----GGLVYRAGSLIAVLRLQTNNSDQF 119

Db 273 --LIENNMFKA-----INPVSRYDCKNGFWLKNNTI-----TRPAQFSTYSI 316

QY 120 VMNI-----YANNDYVPTG-----CDVSARDVPTLPDYG 152

Db 317 TWTDATKRYVADSWTSIGTPTVAYNTPVSAQCVKDLPGYAG 361

RESULT 36

HEMA_IADU3

ID HEMA_IADU3 STANDARD; PRT; 566 AA.

AC P03442;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain].

GN HA.

OS Influenza A virus (strain A/Duck/Ukraine/1/63).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

CC Influenza A viruses; Influenzavirus A.

OX NCBI_TaxID=11374;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=8202542; PubMed=6169439;

RA Fang R., Min Jou W., Huylebroeck D., Devos R., Plets W.;

RT "Complete structure of A/duck/Ukraine/63 influenza hemagglutinin gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza hemagglutinin.";

RL Cell 25:315-323(1981).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL: V01087; CAA24271.1; -.

DR PIR: A04053; HMTVDU.

DR HSSP: P03437; 2YIU.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS: PR00329; HEMAGGLUTN12.

DR ProDom: PD000225; Hemagglutn.1.

DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.

FT SIGNAL 1 16

FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.

FT CARBOHYD 346 366 HEMAGGLUTININ HA2 CHAIN.

FT CARBOHYD 24 24

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).

SO SEQUENCE 566 AA; 63530 MW; E70B87F0AE1178F4 CRC64;

Query Match 8.8%; Score 74.5; DB 1; Length 566;

Best Local Similarity 25.7%; Pred. No. 18;

Matches 44; Conservative 18; Mismatches 70; Indels 39; Gaps 8;

QY 1 PVVAVGQNLVVDL---STQIFCHNDYPTITDYVTLGRGSAYGVLN-----NFSQTVK 51

| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

DB 90 PHCDVFQNETWDLFEVRSNAF-SNCPYDIPDYASLRSLVASSGTLTEFTGFWTGVQ 148
QY 52 YGSSYPPTTSEPRVYVNSRTDKPVALYLTLPVSSAGGLVKGSLAVILRTQTN 111
DB 149 NGSSS-----ACKRGAPANGFESRLN-W-----LTRSSAVPLVAVMPNN 187
QY 112 YNSDDQFVNNIY-----ANNDDVYVPTGCDVSARDYVTLTPDYKGSVP 155
DB 188 DNDKDL-YINGVHHPTNQGTLNLYOGASGRVYVSTRRSQGTIIIPNIGSNP 237
RESULT 37
POLG_EC11G STANDARD; PRT; 2195 AA.
ID P29813: 066785.
DT 01-APR-1993 (Rel. 25, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polypeptide [Contains: Coat protein VP4 (P1A); Coat protein VP2 (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein P3A; Genome-linked protein VP6 (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
DE Echovirus 11 (strain Gregory).
OS Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae; Enterovirus.
OC NCBI_TaxID=31705;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95282506; PubMed=7762294;
RA Dehliand L., Nissinen L., Pulli T., Hytinen V.P., Stanway G., Hyypia T.;
RT "The genome of echovirus 11."
RL Virus Res. 35:215-222(1995).
RN [2]
RP SEQUENCE OF 822-2195 FROM N.A.
RX MEDLINE=91011360; PubMed=2170575;
RA Avineni P., Hyypia T.;
RT "Echoviruses include genetically distinct serotypes."
RL J. Gen. Virol. 71:2133-2139(1990).
CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE AT CERTAIN O/G SITES IN THE POLYPEPTIDE. THEY ARE CYSTEINE PROTEASES.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-Gly bond in the poliovirus polypeptide. In other picornavirus reactions Gln may be substituted for Gln, and Ser or Thr for Gly.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-Gly bond in the picornavirus polypeptide. In other picornavirus reactions Gln may be substituted for Gln, and Ser or Thr for Gly.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate + (RNA)(N).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS. CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/P2A IS CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X80059; CA56365.1; -
DR EMBL: D10582; BAA01439.1; -
DR PIR: A36642; GNYPEC.
DR HSP: P21404; IDAM.
DR MEROPS: C03.011; -.

DR MEROPS: C03.022; -
DR InterPro: IPR004004; Calici_pol_hel.
DR InterPro: IPR000199; Cys-protease-3c.
DR InterPro: IPR003138; Pico_P1A.
DR InterPro: IPR000081; Pico_P2A.
DR InterPro: IPR002527; Pico_P2B.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR InterPro: IPR001676; Rnv.
DR Pfam: PR00073; rnv. 3.
DR Pfam: PF00548; Cys-protease-3c; 1.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR PRINTS: PR00918; CALICVIRUSNS.
DR ProDom: PD001125; Cys-protease-3c; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR ProDom: PD001306; Pico_P2A; 1.
KW Polypeptide; Coat protein; Core protein; Core protein; Transferase; Myristate; RNA-directed RNA polymerase; Hydrolyase; Thiol protease.
FT CHAIN 2
FT CHAIN 69
FT CHAIN 70
FT CHAIN 331
FT CHAIN 332
FT CHAIN 569
FT CHAIN 570
FT CHAIN 861
FT CHAIN 862
FT CHAIN 1011
FT CHAIN 1012
FT CHAIN 1110
FT CHAIN 1111
FT CHAIN 1439
FT CHAIN 1440
FT CHAIN 1528
FT CHAIN 1529
FT CHAIN 1550
FT CHAIN 1551
FT CHAIN 1733
FT CHAIN 1734
FT CHAIN 2195
FT LIPID 2
FT ACT_SITE 1697
FT ACT_SITE 1697
FT ACT_SITE 1711
FT ACT_SITE 1711
FT ACT_SITE 823
FT ACT_SITE 827
SQ SEQUENCE 2195 AA; 245405 MW; 1CFCSDFE28831AF0 CRC64;
Query Match 8.8%; Score 74.5; DB 1; Length 2195;
Best Local Similarity 21.5%; Pred. No. 92;
Matches 39; Conservative 24; Mismatches 61; Indels 57; Gaps 9;
QY 1 PVVWVGQNLVVDLSTQIFCHNDYPERITDYVTLVLRGSAVGVLSNFSGT--VKYSGSSYP 58
DB 753 PFLSIG-----NAYSNFVDGMSHFQNGVYGVNTLNHMGQYVRHVNSSP 798
QY 59 FPTTSEPRVYVNSRTDKP-----PVALYLP----- 86
DB 799 LPMTS-TVRMYFKRHKAVVPRPRLOCYKMASTVFPTNVTDKRTSINYIPETVKPD 857
QY 87 VSSAGGLVIRAGSLIIVLIRQTNVY--NSDDQFVNNIYANNVVPVPT---GGCDVSA 140
DB 858 LSNYGARFGYSGA-VYVYVNRVVRHLATRDMONCVMEDI--NDLLISTTAAHCYVIA 915
QY 141 R 141
DB 916 R 916
RESULT 38
HEMA_IAEN6 STANDARD; PRT; 328 AA.
ID HEMA_IAEN6
AC P04664;
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/England/878/69).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11377;

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=81194918; PubMed=6164798;
RX Sleigh M.J., Both G.W., Underwood P.A., Bender V.J.;
RA "Attenuate drift in the hemagglutinin of the Hong Kong influenza
RT subtype: correlation of amino acid changes with alterations in viral
RL antigenicity.";
RL J. Virol. 37:845-853(1981);
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K03335; AAA43184.1; -
DR HSSP: P03437; 1HGE.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT CARBHYD 1 328
FT CHAIN 1
FT NON_TER 1
FT CARBOHYD 8
FT CARBOHYD 22
FT CARBOHYD 38
FT CARBOHYD 63
FT CARBOHYD 165
FT CARBOHYD 285
FT NON_TER 328
SQ SEQUENCE 328 AA; 36072 MW; 9C3A86B3A8D856FE CRC64;
Query Match 8.7%; Score 73.5; DB 1; Length 328;
Best Local Similarity 24.3%; Pred. No. 12;
Matches 36; Conservative 19; Mismatches 58; Indels 35; Gaps 7;
QY 21 NDYETTDVYTLORGSAVGVLS-----NFGTVKYSGSSYPPTSETPRVYNSRT 74
DB 96 NCYDYDVPDYSRLVASSGTEFTEGFTWGTGVTONGSN---ACKRGPDGFSRL 151
QY 75 DKPPVALYLTPVSSAGGLVYKAGSLAVILRLTNNVNSDDPQFVNNIY---ANND--- 128
DB 152 N-W-----LTKSGSTYPLNVTPMNNDFDKL-YVGVHHPSTNOETS 193
QY 129 -VVVPTGCDVSARDYVTLPLDYGSRV 155
DB 194 LYYQASGRVYVSTRSQTITIPNIGSRP 221
RESULT 39
NK3R_RAT
ID NK3R_RAT STANDARD; PRT; 452 AA.
AC P16177;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Neuromedin K receptor (NK3R) (Neurokinin B receptor) (NK-3 receptor)
DE (NK-3R).
GN TACR3 OR TACR3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;

RX MEDLINE=90110113; PubMed=2153106;
RA Shigemoto R., Yokota Y., Tsuchida K., Nakanishi S.;
RT "Cloning and expression of a rat neuromedin K receptor cDNA.";
RL J. Biol. Chem. 265:623-628(1990).
CC -1- FUNCTION: THIS IS A RECEPTOR FOR THE TACHIKININ NEUROPEPTIDE
CC NEUROMEDIN K (NEUROKININ B). IT IS ASSOCIATED WITH G PROTEINS
CC THAT ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER
CC SYSTEM.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- PM: THE ANCHORING OF THIS RECEPTOR TO THE PLASMA MEMBRANE IS
CC PROBABLY MEDIATED BY THE PALMITOYLATION OF A CYSTEINE RESIDUE.
CC -1- MISCELLANEOUS: THE RANK ORDER OF AFFINITY OF THIS RECEPTOR TO
CC TACHIKININS IS: NEUROMEDIN K > SUBSTANCE P.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC HIGHEST TO OTHER TACHIKININS RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J05189; AAA41688.1; -
DR PIR: A34916; A34916.
DR HSSP: P02699; 1P88.
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCR_Rhodopsn.
DR PROSITE: PS00237; G-PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS50262; G-PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Phosphorylation; Lipoprotein; Palmitate.
FT DOMAIN 1 71
FT TRANSMEM 72 94
FT DOMAIN 95 104
FT TRANSMEM 105 126
FT DOMAIN 127 146
FT TRANSMEM 147 168
FT DOMAIN 169 188
FT TRANSMEM 189 209
FT DOMAIN 210 232
FT TRANSMEM 233 257
FT DOMAIN 258 286
FT TRANSMEM 287 308
FT DOMAIN 309 321
FT TRANSMEM 322 346
FT DOMAIN 347 452
FT CARBOHYD 9
FT CARBOHYD 23
FT CARBOHYD 40
FT CARBOHYD 60
FT DISULFID 145
FT LIPID 361
SQ SEQUENCE 452 AA; 51106 MW; 51D8FC94F34CAE CRC64;
Query Match 8.6%; Score 73; DB 1; Length 452;
Best Local Similarity 28.3%; Pred. No. 19;
Matches 28; Conservative 16; Mismatches 23; Indels 33; Gaps 6;
QY 27 ITDYVTLORGSAVGVLSNFGTVKYSGSSYPPTSETPRVYNSRTDK---PPVAL 82
DB 30 VTETMLAQAG-----NFS-----SALGIPATTPAQSPVQRANITNOFVPSMRIAL 74
QY 83 YLTPVSSAGGLVYKAGSLAVILRLTNNVNSDDPQFVNNIY---ANND--- 112
DB 75 W-----SLAYGVAVAVVAGFNLIVITIIIAHKRRRTVNTY 109
RESULT 40
FLGE_CAUCR
ID FLGE_CAUCR STANDARD; PRT; 591 AA.


```

RESULT 42
GABB_HUMAN
ID GABB_HUMAN STANDARD: PRT: 383 AA.
AC 006547.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GA binding protein beta-1 chain (GABP-beta-1 subunit) (Transcription
factor E4F1-53) (GABPB1) (Nuclear respiratory factor-2 subunit beta
2).
GN GABPB1 OR GABPB OR E4F1B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 1.
RP SEQUENCE FROM N.A.
RX MEDLINE=93180783; PubMed=8441384;
RA Matanabe H., Sawada J.-I., Yano K.-I., Yamaguchi K., Goto M.,
Handa H.;
RT "cDNA cloning of transcription factor E4F1 subunits with Ets and
notch motifs.";
RL Mol. Cell. Biol. 13:1385-1391(1993).
RN 12.
RP SEQUENCE FROM N.A.
RX MEDLINE=95097980; PubMed=7799916;
RA Guenaga S., Virasatus J.V., Scarpulla R.C.;
RT "Four structurally distinct, non-DNA-binding subunits of human
nuclear respiratory factor 2 share a conserved transcriptional
activation domain.";
RL Mol. Cell. Biol. 15:102-111(1995).
CC -1- FUNCTION: TRANSCRIPTION FACTOR CAPABLE OF INTERACTING WITH PURINE
RICH REPEATS (GA REPEATS). NECESSARY FOR THE EXPRESSION OF THE
ADENOVIRUS E4 GENE.
CC -1- SUBUNIT: HETEROTETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; beta-1 (shown here) and beta-2
(NC 006545); are produced by alternative splicing.
CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL, D13317, BAA02574.1; +
DR EMBL, U13046; AAA65708.1; +
DR HSSP; Q00420; IAWC.
DR TRASNFPAC; T01391; +
DR Genew; HGNC:4073; GABPB1.
DR MIM; 600610; +
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 4.
DR PRINTS; PRO1415; ANKYRN.
DR SMART; SM00248; ANK; 3.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
KW Transcription regulation; Nuclear protein; Alternative splicing;
KW ANK repeat; Repeat.
FT REPEAT 5 34 ANK 1.
FT REPEAT 37 66 ANK 2.
FT REPEAT 70 99 ANK 3.
FT REPEAT 103 132 ANK 4.
FT REPEAT 136 166 ANK 5.
SQ SEQUENCE 383 AA; 41321 MW; 07E7081A60016288 CRC64;
Query Match 8.5%; Score 72.5; DB 1; Length 383;
Best Local Similarity 20.0%; Pred. No. 18;
Matches 36; Conservative 36; Mismatches 53; Indels 53; Gaps 9;

```

```

QY 4 NVGQNLVVDLSQIFCHNDYPETITDYY-----TLQGSAYGG 41
DB 151 DIAELIQAAMONQINFTNPEPTVITIAATPOFIIGEGVYNLDETCVSAVQFNSSTS 210
QY 42 VLSNFGTVKYSGGSSYPPTTSETPRVVYNSRTDKPMFVALYLPVSSAGCL--VIRKGS 99
DB 211 VLATIALALAE--ASAPLSNSESPPVVA-----TTE-----VVAESVDGAIQGVSSGG 257
QY 100 LIAVILLIQTNNYNSDDPQFWMNIYANNDDVYPPVPGCCVSAKDVTYVLPIDRGSVPIPLT 159
DB 258 QQVITIV-----TDGIQ-LGNLHS-----IPISGI---GQPIIVMPDQGVLTVPAT 301

RESULT 43
FLA2_PYRKO
ID FLA2_PYRKO STANDARD: PRT; 580 AA.
AC Q9V2X0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Flagellin B2 precursor.
GN FLAB2.
OS Pyrococcus kodakarensis.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Thermococcus.
OX NCBI_TaxID=69014;
RN 1.
RP SEQUENCE FROM N.A.
RX STRAIN=KOD1;
RX MEDLINE=99413236; PubMed=10483738;
RA Nagahisa K., Ezaki S., Fujiwara S., Imanaka T., Takagi M.;
RT "Sequence and transcriptional studies of five clustered flagellin
genes from hyperthermophilic archaeon Pyrococcus kodakarensis KOD1.";
RL FEMS Microbiol. Lett. 178:183-190(1999).
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
FORM THE FILAMENTS OF FLAGELLA.
CC -1- SIMILARITY: BELONGS TO THE ARCHAEAL FLAGELLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AB018434; BAA84106.1; +
DR InterPro; IPR002774; Arch_Flagellin.
DR Pfam; PF01917; Arch_Flagellin; 1.
KW Flagella; Multigene family.
FT PROPEP 1 9 BY SIMILARITY.
FT CHAIN 10 580 FLAGELLIN B2.
SQ SEQUENCE 580 AA; 61138 MW; AC92D97B1C896A2B CRC64;
Query Match 8.5%; Score 72.5; DB 1; Length 580;
Best Local Similarity 25.4%; Pred. No. 29;
Matches 32; Conservative 19; Mismatches 42; Indels 33; Gaps 5;
QY 42 VLSNFGTVKYSGGSSYPPTTSEPT-----RVVYNSRTDKPMF-----VALITPVSAA 90
DB 32 VLINTSYLQOKASTGRETTQEVASGLIKMKYIGYDPADPPASGKITERLAYVYVSPAGS 91
QY 91 GGLVIRKAGSLIAVLILQNTNNYNSDDPQFWMNIYANNDDVYPPVPGCCVSAKDVTYVLPID 147
DB 92 SGIDMKKVRVY-----LSMGDKQALINYI-----VPESGTVSETPTTIKLAF 135
QY 148 ---PDY 150
DB 136 TSEPDW 141
RESULT 44
RUVB_RICPR

```



```

ID  RVB_RICPR      STANDARD:      PRG:      342 AA.
AC  092DE5;
DT  30-MAY-2000 (Rel. 39, Created)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Holiday junction DNA helicase RVB.
CN  RVB OR RP386.
OS  Rickettsia prowazekii.
OC  Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC  Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX  NCBI_TaxID=782;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Madrid E;
RX  MEDLINE=99039499; PubMed=9823893;
RA  Anderson S.G.E., Zomorodipour A., Andersson J.O.,
RA  Slicheritz-Polten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA  Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT  "The genome sequence of Rickettsia prowazekii and the origin of
RT  mitochondria."
RL  Nature 396:133-140(1998).
CC  -1- FUNCTION: THE RVB-RVB COMPLEX IN THE PRESENCE OF ATP REMOVES
CC  CRUCIFORM STRUCTURE IN SUPERCOILED DNA WITH PALINDROMIC SEQUENCE,
CC  INDICATING THAT IT MAY PROMOTE STRAND EXCHANGE REACTIONS IN
CC  HOMOLOGOUS RECOMBINATION. RVB IS AN HELICASE THAT MEDIATES THE
CC  HOLIDAY JUNCTION MIGRATION BY LOCALIZED DENATURATION AND
CC  REANNEALING (BY SIMILARITY).
CC  -1- SUBUNIT: PARTICIPATES TO UV-TOLERANCE OF SYNECHOCYSTIS PCC 6803.
CC  -1- SUBUNIT: FORMS A COMPLEX WITH RVB.
CC  -1- SIMILARITY: BELONGS TO THE RVB FAMILY.
CC  -----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: AJ235271; CAI14843.1;
DR  Interpro: IPR003593; AAA_ATPase.
DR  Interpro: IPR003959; AAA_ATPase_central.
DR  Interpro: IPR004605; RVB.
DR  Pfam: PF00004; AAA.1.
DR  SMART: SM00382; AAA.1.
DR  TIGRfams: TIGR00635; RVB.1.
KW  DNA repair; SOS response; ATP-binding; DNA recombination; Helicase;
KW  Complete proteome.
FT  NP_BIND 57 64 ATP (POTENTIAL)
SQ  SEQUENCE 342 AA; 38435 MW; D2F9290E84079BA5 CRC64;

Query Match 8.5%; Score 72; DB 1; Length 342;
Best Local Similarity 23.1%; Pred. No. 17;
Matches 25; Conservative 15; Mismatches 26; Indels 42; Gaps 4;

QY 88 SSAGGLVIRKAGSLIVLILQTNNNYNSDDPOFVNIYAN-----DVVV 131
   ||| ||| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 80 STAGPAITKADLASIL-----TNLEKNDVLFIDEIHLTLVEVLYSAMDEFDITII 134
   ||| ||| |::| |::| |::| |::| |::| |::| |::| |::| |::|

QY 132 PTGGCDVSARDVTVTLDPY-----RGSVPILPLVY 161
   ||| ||| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 135 GEGS---AARPVKITLPKFLIGATTREFGLISNPLDRDRGIPRLNMFY 179
   ||| ||| |::| |::| |::| |::| |::| |::| |::| |::| |::|

RESULT 45
HEMA_IADHM
ID  HEMA_IADHM      STANDARD:      PRG:      550 AA.
AC  P43259;
DT  01-NOV-1995 (Rel. 32, Created)
DT  01-NOV-1995 (Rel. 32, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE  Hemagglutinin HA2 chain] (Fragment).

```

```

GN  HA
OS  Influenza A virus (strain A/Duck/Hong Kong/231/77).
OC  Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC  Influenza A viruses; Influenzavirus A.
OX  NCBI_TaxID=45411;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=91341491; PubMed=1875195;
RA  Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT  "Molecular evidence for a role of domestic ducks in the introduction
RT  of avian H3 influenza viruses to pigs in southern China, where the
RT  A/Hong Kong/68 (H3N2) strain emerged."
RL  J. Gen. Virol. 72:2007-2010(1991).
CC  -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC  CELL RECEPTORS AND FOR INITIATING INFECTION.
CC  -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC  (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC  -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC  -----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: D00932; BAA00772.1;
DR  HSSP: P03437; 2V10.
DR  Interpro: IPR001364; Hemagglutn.
DR  Pfam: PF00509; Hemagglutinin.1.
DR  Prodom: PD000225; Hemagglutn.1.
KW  Envelope protein; Hemagglutinin; Glycoprotein.
FT  NON_TER 1 1
FT  CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT  CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.
FT  CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ  SEQUENCE 550 AA; 61762 MW; 6FEF8B49488C191A CRC64;

Query Match 8.5%; Score 72; DB 1; Length 550;
Best Local Similarity 22.8%; Pred. No. 30;
Matches 36; Conservative 18; Mismatches 28; Indels 76; Gaps 8;

QY 21 NDYPTITDYTLQ-----RGSAYGVLSN-----FS--GVK 51
   ||| ||| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 96 NCPYDVPDYASIRSLVASSGTLFTEGFVWGVYNGGSMACKRGPRANGFSRLNMLT 155
   ||| ||| |::| |::| |::| |::| |::| |::| |::| |::| |::|

QY 52 YGSSSP-----PPTTS-----TPRVYNSPT----- 74
   ||| ||| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 156 KSGSTIPVLNTPMNNNDKLYIWMGIIHPSDEQETLYVOASGRVIVSTRSOQTII 215
   ||| ||| |::| |::| |::| |::| |::| |::| |::| |::| |::|

QY 75 ---DKFW-----PVALLTPVSSAGLVYKA--GSLIA 102
   ||| ||| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 216 NIGSRPWVWGSGRISITVITVYKSGDVLYINSNGNLIA 253
   ||| ||| |::| |::| |::| |::| |::| |::| |::| |::| |::|

Search completed: November 28, 2002, 17:20:43
Job time : 28 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 10:37:10 ; Search time 26 seconds
(without alignments)
1275.907 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186

Perfect score: 848
Sequence: 1 PVVNVGNLVLDSTQIFCH.....DVTYTLDPYRGSPVPIPLTVY 161

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match length	DB ID	Description
1	834	98.3	300	2	Q9S497
2	834	98.3	300	2	Q9S496
3	828	97.6	300	2	Q9S494
4	828	97.6	300	2	Q9R512
5	827	97.5	300	2	Q9F551
6	826	97.4	300	2	Q9F550
7	824	97.2	300	16	Q8XBA6
8	823	97.1	300	2	Q9S495
9	822	96.9	300	2	Q9F627
10	819	96.6	300	2	Q87634
11	819	96.6	300	2	Q9AP05
12	816	96.2	300	2	Q9S6R1
13	756	89.2	243	2	Q08858
14	618	72.9	166	2	Q9F5R9
15	384	45.3	304	16	Q8XAX2
16	262	30.9	129	2	Q9XBV8

17	207	24.4	299	2	Q53298	Q53298 escherichia
18	199.5	23.5	299	2	Q46686	Q46686 escherichia
19	89	10.5	684	16	Q911P8	Q911P8 streptomyces
20	88.5	10.4	315	12	Q9YN73	Q9YN73 vaccinia vl
21	87	10.3	270	10	Q9YN68	Q9YN68 cowpox viru
22	85.5	10.1	313	12	Q9YN66	Q9YN66 monkeypox v
23	85.5	10.1	313	12	Q8V4S6	Q8V4S6 monkeypox v
24	84	9.9	1755	17	Q26812	Q26812 methanobact
25	83.5	9.8	1114	5	Q05352	Q05352 entamoeba h
26	82.5	9.7	315	12	Q9JF47	Q9JF47 vaccinia vl
27	82.5	9.7	1155	17	Q58048	Q58048 pyrococcus
28	82	9.7	383	3	Q9X6L5	Q9X6L5 klebsiella
29	81.5	9.6	315	12	Q89120	Q89120 vaccinia vl
30	81.5	9.6	566	12	Q82496	Q82496 influenza vi
31	81	9.6	824	10	Q40713	Q40713 oryza sativ
32	80.5	9.5	473	10	Q9SE00	Q9SE00 ipomoea bat
33	80.5	9.5	1304	2	Q9XC46	Q9XC46 rickettsia
34	80.5	9.5	1620	2	Q9XKB5	Q9XKB5 rickettsia
35	80	9.4	270	12	Q9YN72	Q9YN72 vaccinia vl
36	80	9.4	519	3	Q96UK8	Q96UK8 tremetes ve
37	79.5	9.4	315	12	Q89182	Q89182 vaccinia vl
38	79.5	9.4	566	12	Q67126	Q67126 influenza vi
39	79.5	9.4	1924	12	Q9JF52	Q9JF52 ectromelia
40	79	9.3	421	9	Q64302	Q64302 bacterioph
41	79	9.3	985	5	Q9YFW7	Q9YFW7 dirosophila
42	79	9.3	1064	10	Q9FX19	Q9FX19 arabidopsis
43	79	9.3	3763	5	Q8T2A1	Q8T2A1 dictyostell
44	78.5	9.3	398	16	Q8U9U1	Q8U9U1 agrobacteri
45	78.5	9.3	782	2	Q93SH4	Q93SH4 bradyrhizob

ALIGNMENTS

RESULT 1	Q9S497	PRELIMINARY;	PRT;	300 AA.
ID	Q9S497			
AC	Q9S497			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	FlmH.			
GN	FlmH.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-K12;			
RX	MEDLINE-20327582; PubMed-10869080;			
RA	Hemrick T.S., Harris S.L., Spears P.A., Havelle E.A., Horton J.R.,			
RA	Russell P.W., Orndorff P.E.,			
RT	"Genetic characterization of Escherichia coli type 1 pilus adhesin			
RT	mutants and identification of a novel binding phenotype.";			
RL	J. Bacteriol. 182:4012-4021(2000).			
DR	EMBL; AF154925; AAD4319.1; -			
DR	InterPro: IPR000259; Fimbrin.			
DR	Pfam: PF00419; Fimbrin; 1.			
FT	SEQUENCE 300 AA; 31517 MW; D29340A1575A2358 CRC64;			
FT	VARIANT 19 19 D -> A.			

Query Match 98.3% Score 834; DB 2; Length 300;

Best Local Similarity 98.1% Pred. No. 4e+71;

Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY	1	PVVNVGNLVLDSTQIFCHNDYPTDYTLDRGSAVGVSNFSGTYKYSSTPP	60
DB	47	PVVNVGNLVLDSTQIFCHNDYPTDYTLDRGSAVGVSNFSGTYKYSSTPP	106
QY	61	TTSTPRVYNSRFDKMPVAVLYLPVSSAGGLVKGSLIAVILLOTNNYNSDDPQV	120
DB	107	TTSTPRVYNSRFDKMPVAVLYLPVSSAGGLVKGSLIAVILLOTNNYNSDDPQV	166

```
QY 121 WNIYANDVVPPTGGCDVSARDVTTLTPDYGSPVPIPLTVY 161
    |||||||
DB 167 WNIYANDVVPPTGGCDVSARDVTTLTPDYGSPVPIPLTVY 207

RESULT 2
Q9S496 PRELIMINARY: PRT: 300 AA.
ID Q9S496;
AC Q9S496;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE FIMH.
GN Escherichia coli.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=20327582; PubMed=10869080;
RA Hamrick T.S., Harris S.L., Spears P.A., Havelle E.A., Horton J.R.,
  Russell P.W., Orndorff P.E.;
RT "Genetic characterization of Escherichia coli type 1 pilus adhesin
  mutants and identification of a novel binding phenotype.";
RL J. Bacteriol. 182:4012-4021(2000).
DR EMBL: AF154926; AAD4320.1;
DR InterPro: IPR000259; Fimbr1al.
DR Pfam: PF00419; Fimbr1al; 1.
FT VARIANT 21
SQ SEQUENCE 300 AA; 31499 MW; 3ADF97A85F8CE478 CRC64;

Query Match 98.3%; Score 834; DB 2; Length 300;
Best Local Similarity 98.1%; Pred. No. 4e-71;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PNVNNGQLVVDLSTQIFCHNDYPETITDYVTLQSGAAYGVLNFSGTVKYSGSSYPFP 60
    |||||||
DB 47 PNVNNGQLVVDLSTQIFCHNDYPETITDYVTLQSGAAYGVLNFSGTVKYSGSSYPFP 106

QY 61 TTSETPRVVNSRTDKPMPVALYLTLPVSSAGGLVIRAGSLIAVLILRQTNVNSDDFOFV 120
    |||||||
DB 107 TTSETPRVVNSRTDKPMPVALYLTLPVSSAGGLVIRAGSLIAVLILRQTNVNSDDFOFV 166

QY 121 WNIYANDVVPPTGGCDVSARDVTTLTPDYGSPVPIPLTVY 161
    |||||||
DB 167 WNIYANDVVPPTGGCDVSARDVTTLTPDYGSPVPIPLTVY 207

RESULT 3
Q9S494 PRELIMINARY: PRT: 300 AA.
ID Q9S494;
AC Q9S494;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE FIMH.
GN Escherichia coli.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=20327582; PubMed=10869080;
RA Hamrick T.S., Harris S.L., Spears P.A., Havelle E.A., Horton J.R.,
  Russell P.W., Orndorff P.E.;
RT "Genetic characterization of Escherichia coli type 1 pilus adhesin
  mutants and identification of a novel binding phenotype.";
RL J. Bacteriol. 182:4012-4021(2000).
```

```
DR EMBL: AF154928; AAD4322.1;
DR InterPro: IPR000259; Fimbr1al.
DR Pfam: PF00419; Fimbr1al; 1.
FT VARIANT 79
SQ SEQUENCE 300 AA; 31516 MW; B6073DF6836B86E CRC64;

Query Match 97.6%; Score 828; DB 2; Length 300;
Best Local Similarity 97.5%; Pred. No. 1.5e-70;
Matches 157; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 PNVNNGQLVVDLSTQIFCHNDYPETITDYVTLQSGAAYGVLNFSGTVKYSGSSYPFP 60
    |||||||
DB 47 PNVNNGQLVVDLSTQIFCHNDYPETITDYVTLQSGAAYGVLNFSGTVKYSGSSYPFP 106

QY 61 TTSETPRVVNSRTDKPMPVALYLTLPVSSAGGLVIRAGSLIAVLILRQTNVNSDDFOFV 120
    |||||||
DB 107 TTSETPRVVNSRTDKPMPVALYLTLPVSSAGGLVIRAGSLIAVLILRQTNVNSDDFOFV 166

QY 121 WNIYANDVVPPTGGCDVSARDVTTLTPDYGSPVPIPLTVY 161
    |||||||
DB 167 WNIYANDVVPPTGGCDVSARDVTTLTPDYGSPVPIPLTVY 207

RESULT 4
Q9RSY2 PRELIMINARY: PRT: 300 AA.
ID Q9RSY2;
AC Q9RSY2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE FIMH SUBUNIT=MANNOSSE-sensitive type 1 fimbrial adhesin.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94131954; PubMed=7905476;
RA Sokuneko E.V., Courtney H.S., Ohman D.E., Klemm P., Hasly D.L.;
RL J. Bacteriol. 176:748-753(1994).
DR InterPro: IPR000259; Fimbr1al.
DR Pfam: PF00419; Fimbr1al; 1.
SQ SEQUENCE 300 AA; 31552 MW; E75328B7C366A319 CRC64;

Query Match 97.6%; Score 828; DB 2; Length 300;
Best Local Similarity 97.5%; Pred. No. 1.5e-70;
Matches 157; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 PNVNNGQLVVDLSTQIFCHNDYPETITDYVTLQSGAAYGVLNFSGTVKYSGSSYPFP 60
    |||||||
DB 47 PNVNNGQLVVDLSTQIFCHNDYPETITDYVTLQSGAAYGVLNFSGTVKYSGSSYPFP 106

QY 61 TTSETPRVVNSRTDKPMPVALYLTLPVSSAGGLVIRAGSLIAVLILRQTNVNSDDFOFV 120
    |||||||
DB 107 TTSETPRVVNSRTDKPMPVALYLTLPVSSAGGLVIRAGSLIAVLILRQTNVNSDDFOFV 166

QY 121 WNIYANDVVPPTGGCDVSARDVTTLTPDYGSPVPIPLTVY 161
    |||||||
DB 167 WNIYANDVVPPTGGCDVSARDVTTLTPDYGSPVPIPLTVY 207

RESULT 5
Q9F5S1 PRELIMINARY: PRT: 300 AA.
ID Q9F5S1;
AC Q9F5S1;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE FIMH.
GN Escherichia coli.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
```

```

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-21289098; PubMed-11395476;
RA Harris S.L., Spears P.A., Haveli E.A., Hamrick T.S., Horton J.R.,
  Orndorff P.E.;
RT "Characterization of Escherichia coli Type 1 Plus Mutants with
  Altered Binding Specificities";
RL J. Bacteriol. 183:4099-4102(2001).
DR EMBL: AF306535; AAG30925.1; -.
DR InterPro: IPR000259; Fimbrin.
DR Pfam: PF00419; Fimbrin; 1.
FT VARIANT 79 79 P -> R.
SQ SEQUENCE 300 AA; 31457 MW; 9F7B0A35FF51F938 CRC64;

Query Match 97.5%; Score 827; DB 2; Length 300;
Best Local Similarity 97.5%; Pred. No. 1.8e-70;
Matches 157; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 PVVAVGNLVVDLSTQIFCHNDYPTTDTVTLQRGSAVGGVLSNFGSTVKYSGSSYFP 60
DB 47 PVVAVGNLVVDLSTQIFCHNDYPTTDTVTLQRGSAVGGVLSNFGSTVKYSGSSYFP 106
OY 61 TTSETPRVYNSRTDKPMPVALYLTTPVSSAGLVKAGSLIAYLILRQTNVNSDDPQFV 120
DB 107 TTSETPRVYNSRTDKPMPVALYLTTPVSSAGLVKAGSLIAYLILRQTNVNSDDPQFV 166
OY 121 WNIYANDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
DB 167 WNIYANDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 207

RESULT 6
OQ9F550 PRELIMINARY; PRT; 300 AA.
ID OQ9F550;
AC OQ9F550;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE FIMH.
GN Escherichia coli.
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-21289098; PubMed-11395476;
RA Harris S.L., Spears P.A., Haveli E.A., Hamrick T.S., Horton J.R.,
  Orndorff P.E.;
RT "Characterization of Escherichia coli Type 1 Plus Mutants with
  Altered Binding Specificities";
RL J. Bacteriol. 183:4099-4102(2001).
DR EMBL: AF306535; AAG30925.1; -.
DR InterPro: IPR000259; Fimbrin.
DR Pfam: PF00419; Fimbrin; 1.
FT VARIANT 163 163 S -> F.
SQ SEQUENCE 300 AA; 31413 MW; 922904051758746D CRC64;

Query Match 97.4%; Score 826; DB 2; Length 300;
Best Local Similarity 97.5%; Pred. No. 2.3e-70;
Matches 157; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 PVVAVGNLVVDLSTQIFCHNDYPTTDTVTLQRGSAVGGVLSNFGSTVKYSGSSYFP 60
DB 47 PVVAVGNLVVDLSTQIFCHNDYPTTDTVTLQRGSAVGGVLSNFGSTVKYSGSSYFP 106
OY 61 TTSETPRVYNSRTDKPMPVALYLTTPVSSAGLVKAGSLIAYLILRQTNVNSDDPQFV 120
DB 107 TTSETPRVYNSRTDKPMPVALYLTTPVSSAGLVKAGSLIAYLILRQTNVNSDDPQFV 166

```

```

OY 121 WNIYANDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
DB 167 WNIYANDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 207

RESULT 7
OQ8XB6 PRELIMINARY; PRT; 300 AA.
ID OQ8XB6;
AC OQ8XB6;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Minor fimbrial subunit, D-mannose specific adhesin.
GN FIMH OR 25918 OR ECSS279.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / ATCC 700927;
RX MEDLINE-21074935; PubMed-11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
  Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
  Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
  Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
  Apodaca J., Anantharaman T.S., Lin J., Yan G., Schwartz D.C.,
  Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE-21156231; PubMed-11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
  Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tode T.,
  Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
  Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
  O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
DR EMBL: AE005662; AAG39502.1; -.
DR EMBL: AP002569; BAB38702.1; -.
DR InterPro: IPR000259; Fimbrin.
DR Pfam: PF00419; Fimbrin; 1.
KW Complete proteome.
SQ SEQUENCE 300 AA; 31459 MW; 68EFE713A3D372CB CRC64;

Query Match 97.2%; Score 824; DB 16; Length 300;
Best Local Similarity 96.9%; Pred. No. 3.5e-70;
Matches 156; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PVVAVGNLVVDLSTQIFCHNDYPTTDTVTLQRGSAVGGVLSNFGSTVKYSGSSYFP 60
DB 47 PVVAVGNLVVDLSTQIFCHNDYPTTDTVTLQRGSAVGGVLSNFGSTVKYSGSSYFP 106
OY 61 TTSETPRVYNSRTDKPMPVALYLTTPVSSAGLVKAGSLIAYLILRQTNVNSDDPQFV 120
DB 107 TTSETPRVYNSRTDKPMPVALYLTTPVSSAGLVKAGSLIAYLILRQTNVNSDDPQFV 166
OY 121 WNIYANDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
DB 167 WNIYANDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 207

RESULT 8
OQ9S495 PRELIMINARY; PRT; 300 AA.
ID OQ9S495;
AC OQ9S495;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE FIMH.

```


Best Local Similarity 95.7%; Pred. No. 1.1e-69;
Matches 154; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAVVNGONLVVDSTQIFCHNDYPERITDYVTLQKRSAGGVLSNFGTYKYSGSSYPP 60
Db 50 PAVVNGONLVVDSTQIFCHNDYPERITDYVTLQKRSAGGVLSNFGTYKYSGSSYPP 109
QY 61 TTSETPRVYVNSRTDKPMPALYLTVPVSSAGGLVIRKAGSLIALLIROTNNYNSDDFOFY 120
Db 110 TTSETPRVYVNSRTDKPMPALYLTVPVSSAGGLVIRKAGSLIALLIROTNNYNSDDFOFY 169
QY 121 WNIYANDVVPVPGGCDVSARDVTVTLPDYRGSVPILPLY 161
Db 170 WNIYANDVVPVPGGCDVSARDVTVTLPDYRGSVPILPLY 210

RESULT 12

Q956R1 PRELIMINARY; PRT; 300 AA.

AC Q956R1; 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Type 1 fimbriae adhesin, precursor polypeptide.
GN FTMH.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT78;
RX MEDLINE=88038337; PubMed=2890081;
RA Klemm P., Christiansen G.;
RT "Three fim genes required for the regulation of length and mediation of adhesion of Escherichia coli type 1 fimbriae.";
RL Mol. Gen. Genet. 208:439-445(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MT78;
RX MEDLINE=96235989; PubMed=8636962;
RA Marc D., Dho-Moulin M.;
RT "Analysis of the fim cluster of an avian O2 strain of Escherichia coli: serogroup-specific sites within fima and nucleotide sequence of fimI.";
RL J. Med. Microbiol. 44:444-452(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-MT78;
RX MEDLINE=96125229; PubMed=8550444;
RA Klemm P., Tong S., Nielsen H., Conway T.;
RT "The gntP gene of Escherichia coli involved in gluconate uptake.";
RL J. Bacteriol. 178:61-67(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-MT78;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G.II, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-MT78;
RX MEDLINE=98439336; PubMed=9766199;
RA Marc D., Arne P., Bree A., Dho-Moulin M.;
RT "Colonization ability and pathogenic properties of a fim- mutant of an avian strain of Escherichia coli.";
RL Res. Microbiol. 149:473-485(1998).
DR EMBL: AJ225176; CAI2423.1; InterPro: IPR000259; FimDial.

DR Pfam: PF00419; FimDial: 1.
SQ SEQUENCE 300 AA; 31459 MM; EA040CEDD2137PE8 CRC64;

Query Match 96.2%; Score 816; DB 2; Length 300;
Best Local Similarity 95.7%; Pred. No. 2e-69;
Matches 154; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAVVNGONLVVDSTQIFCHNDYPERITDYVTLQKRSAGGVLSNFGTYKYSGSSYPP 60
Db 47 PAVVNGONLVVDSTQIFCHNDYPERITDYVTLQKRSAGGVLSNFGTYKYSGSSYPP 106
QY 61 TTSETPRVYVNSRTDKPMPALYLTVPVSSAGGLVIRKAGSLIALLIROTNNYNSDDFOFY 120
Db 107 TTSETPRVYVNSRTDKPMPALYLTVPVSSAGGLVIRKAGSLIALLIROTNNYNSDDFOFY 166
QY 121 WNIYANDVVPVPGGCDVSARDVTVTLPDYRGSVPILPLY 161
Db 167 WNIYANDVVPVPGGCDVSARDVTVTLPDYRGSVPILPLY 207

RESULT 13

Q08858 PRELIMINARY; PRT; 243 AA.

AC Q08858; 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMblrel. 08, Last annotation update)
DE FIMH protein precursor.
GN FTMH.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89155420; PubMed=2563996;
RA Gerlach G.F., Clegg S., Allen B.L.;
RT "Identification and characterization of the genes encoding the type 3 and type 1 fimbrial adhesins of Klebsiella pneumoniae.";
RL J. Bacteriol. 171:1262-1270(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA Nichols N.N., Clegg S.;
RL Submitted (Aug-1993) to the EMBL/Genbank/DBI databases.
CC - FUNCTION: INVOLVED IN REGULATION OF LENGTH AND MEDIATION OF ADHESION OF TYPE 1 FIMBRIAE (BUT NOT NECESSARY FOR THE PRODUCTION OF FIMBRIAE).
CC EMBL: I23111; AAA25063.1;
DR FimDial; Signal.
KW FimDial; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 243 FIMH PROTEIN.
SQ SEQUENCE 243 AA; 25977 MM; C8394D/C44A98AE7 CRC64;

Query Match 89.2%; Score 756; DB 2; Length 243;
Best Local Similarity 86.3%; Pred. No. 7.7e-64;
Matches 139; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 1 PAVVNGONLVVDSTQIFCHNDYPERITDYVTLQKRSAGGVLSNFGTYKYSGSSYPP 60
Db 49 PAVVNGONLVVDSTQIFCHNDYPERITDYVTLQKRSAGGVLSNFGTYKYSGSSYPP 108
QY 61 TTSETPRVYVNSRTDKPMPALYLTVPVSSAGGLVIRKAGSLIALLIROTNNYNSDDFOFY 120
Db 109 TTSETPRVYVNSRTDKPMPALYLTVPVSSAGGLVIRKAGSLIALLIROTNNYNSDDFOFY 168
QY 121 WNIYANDVVPVPGGCDVSARDVTVTLPDYRGSVPILPLY 161
Db 169 WNIYANDVVPVPGGCDVSARDVTVTLPDYRGSVPILPLY 209

RESULT 14
Q9F5R9 PRELIMINARY; PRT; 166 AA.

```

AC 09F5R9;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, last annotation update)
DE FIMH.
GN FIMH.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
NCBI_TaxID=562;
RN
  [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA Harris S.L., Spears P.A., Haveli E.A., Hamrick T.S., Horton J.R.,
  Orndorff P.E.;
RT "Isolation and characterization of Escherichia coli type 1 pili
  mutants that have altered binding specificities.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF06537; AAG30927.1; -.
SQ SEQUENCE 166 AA; 17826 MW; 2E64EBA80AE286DC CRC64;

Query Match          72.9%; Score 618; DB 2; Length 166;
Best Local Similarity 98.3%; Pred. No. 5; 9e-51;
Matches 118; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PVVWVGONLVVDSTQIFCHNDYPERITDVTYTLQSGSAYGVLNFGSTVYSGSSYFP 60
DB 47 PVVWVGONLVVDSTQIFCHNDYPERITDVTYTLQSGSAYGVLNFGSTVYSGSSYFP 106

QY 61 TTSETPRVYVNSRTDKPMPVLYLTTPVSSAGGLVYKAGSLAVLILRQTNVNSDDPQFV 120
DB 107 TTSETPRVYVNSRTDKPMPVLYLTTPVSSAGGLVYKAGSLAVLILRQTNVNSDDPQFV 166

RESULT 15
Q8XAX2 PRELIMINARY; PRT; 304 AA.
AC 08XAX2;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-JUN-2002 (TEMBLrel. 21, last sequence update)
DE Putative adhesin, similar to FimH protein.
GN 22206 OR EC82107.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
NCBI_TaxID=83334;
RN
  [1]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
  Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
  Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
  Rasmussen S., Davis N.W., Lim A., Dimallanta E.T., Potamousis K.,
  Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
  Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).

SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
  Han C.-G., Ohtsuda E., Nakayama K., Murata T., Tanaka M., Tobe T.,
  Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
  Khahar S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
  O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL: AE005354; AAG56266.1; -.
DR EMBL: AP002557; BAB35530.1; -.
DR InterPro: IPR000259; Fimbril.
```

```

DR Pfam: PF00419; Fimbril; 1.
KW Complete proteome.
SQ SEQUENCE 304 AA; 32041 MW; EED538023D95AFD5 CRC64;

Query Match          45.3%; Score 384; DB 16; Length 304;
Best Local Similarity 47.9%; Pred. No. 2e-28;
Matches 78; Conservative 29; Mismatches 54; Indels 2; Gaps 2;

QY 1 PVVWVGONLVVDSTQIFCHNDYPERITDVTYTLQSGSAYGVLNFGSTVYSGSSYFP 59
DB 48 PVVWVGONLVVDSTQIFCHNDYPERITDVTYTLQSGSAYGVLNFGSTVYSGSSYFP 107

QY 60 PTSETPRVYVNSRTDKPMPVLYLTTPVSSAGGLVYKAGSLAVLILRQTNVNS-DDPQ 118
DB 108 PTSETPRVYVNSRTDKPMPVLYLTTPVSSAGGLVYKAGSLAVLILRQTNVNS-DDPQ 167

QY 119 FVWNTVANDVYVPTGCDVYARDVYVTLFDYKGVIPILTV 161
DB 168 FVWNTVANDVYVPTGCDVYARDVYVTLFDYKGVIPILTV 210

RESULT 16
Q9XBV8 PRELIMINARY; PRT; 129 AA.
AC 09XBV8;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, last annotation update)
DE FIMH.
GN FIMH.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
NCBI_TaxID=562;
RN
  [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=20327582; PubMed=10869080;
RA Hamrick T.S., Harris S.L., Spears P.A., Haveli E.A., Horton J.R.,
  Russell P.W., Orndorff P.E.;
RT "Genetic characterization of Escherichia coli type 1 plus adhesin
  mutants and identification of a novel binding phenotype.";
RL J. Bacteriol. 182:4012-4021(2000).
DR EMBL: AF154929; AAD44323.1; -.
SQ SEQUENCE 129 AA; 14142 MW; BDE1D4D0E1EC336 CRC64;

Query Match          30.9%; Score 262; DB 2; Length 129;
Best Local Similarity 100.0%; Pred. No. 2.5e-17;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PVVWVGONLVVDSTQIFCHNDYPERITDVTYTLQSGSAYGVLNFGSTV 50
DB 47 PVVWVGONLVVDSTQIFCHNDYPERITDVTYTLQSGSAYGVLNFGSTV 96

RESULT 17
Q53298 PRELIMINARY; PRT; 299 AA.
AC 053298;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, last annotation update)
DE S fimbrial adhesin minor subunit Sfah.
GN SFah.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
NCBI_TaxID=562;
RN
  [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93138776; PubMed=8093693;
RA Hacker J., Kestler H., Hoshutzky H., Jann K., Lottspeich F.,
  Korhonen T.K.;
```


RT "Cloning and characterization of the S fimbrial adhesin II complex of
 RT an *Escherichia coli* O18:K1 meningitis isolate."
 RL Infect. Immun. 61:544-550(1993).
 DR EMBL: S53210; AAB25047.1; -
 DR InterPro: IPR000259; Fimbrlal.
 DR Pfam: PF00419; Fimbrlal; 1.
 SO SEQUENCE 299 AA; 31972 MW; 8EC6B8E84E9067C7 CRC64;

Query Match 24.4%; Score 207; DB 2; Length 299;
 Best Local Similarity 32.6%; Pred. No. 1.2e-11;
 Matches 56; Conservative 22; Mismatches 82; Indels 12; Gaps 5;

OY 1 PVVNGOML-VVDLSTOIFCHN-DYPERITDVTLQSGSAYGVL-----NSFGTVK 51
 DB 34 PVGYDKSTSVLDLSQVSCONEDSTGONYDLKLKSGSPALDPTTYGRDPTSRPT 93
 OY 52 YSGSSYPPTTSERPRVYNSRTDKPMVVALYLPVSSAGGLVTKAGSLIIVLLRQ--T 109
 DB 94 GYARQLPQLQFDQVEARYQYQGWKPPAKLYLPAPGVFKVINGDLATLYVNFST 153
 OY 110 NNYNSDDPQFVWNTYANNDDVYVPTGCGDVASRDVTVTLPTDRGS-VPIPLTV 160
 DB 154 KGOEGERNFTWRYATNDVHIQTCTCRVSSNNKRVLDLPSYPGGVTVPLTV 205

RESULT 18

ID Q46686 PRELIMINARY; PRT; 299 AA.

AC Q46686; Q53398;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE Minor F1652 fimbrial subunit H.
 GN F1652H OR FOCH.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Escherichia*.
 NCBI_TaxID=562;
 RN NCBI
 RP SEQUENCE FROM N.A.
 RC STRAIN=4787 O115:V165:F165;
 RC MEDLINE=95202083; PubMed=7894716;
 RA Haral J., Jacques M., Fairbrother J.M., Bosse M., Forget C.;
 RT "Cloning of determinants encoding F165(2) fimbriae from porcine
 RT septicemic *Escherichia coli* confirms their identity as F1C
 RT fimbriae."
 RL Microbiology 141:221-228(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92073661; PubMed=1683712;
 RA van Die I., Kramer C., Hacker J., Bergmans H., Jongen W., Hoekstra W.;
 RT "Nucleotide sequence of the genes coding for minor fimbrial subunits
 RT of the F1C fimbriae of *Escherichia coli*."
 RL Res. Microbiol. 142:653-658(1991).
 DR EMBL: U09804; AAA74946.1; -
 DR EMBL: S68237; AAB20439.1; -
 DR InterPro: IPR000259; Fimbrlal.
 DR Pfam: PF00419; Fimbrlal; 1.
 SO SEQUENCE 299 AA; 32155 MW; 259C745174C59D05 CRC64;

Query Match 23.5%; Score 199.5; DB 2; Length 299;
 Best Local Similarity 33.3%; Pred. No. 6.1e-11;
 Matches 54; Conservative 24; Mismatches 73; Indels 11; Gaps 6;

OY 10 VVDLSTOIFCHN-DYPERITDVTLQSGSAYGVL-----SNFGTVKSGSSYPPTTS 63
 DB 44 VLDLNLQVLCQNDASGQNDYLRVQGTGCFSPSLAKTYGRDPTNRLSGVSQLPLQO 103
 OY 64 ET-PRVY-NSRTDKPMVVALYLPVSSAGGLVTKAGSLIIVLLRQ--TNNNSDDPQF 119
 DB 104 DTPETAEYQYQGWKPPAKLYLPAPGVFKIILHAGELVAIYVNFSTMGAGEGRNF 163
 OY 120 VNNIYANNDDVYVPTGCGDVASRDVTVTLPTDRGS-VPIPLTV 160

DB 164 TWREYATNDVYIQTCTCRVSSNNKRVLDLPSYPGGVTVPLTV 205

RESULT 19

ID Q9LIF8 PRELIMINARY; PRT; 684 AA.

AC Q9LIF8;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Putative neutral zinc metalloprotease.
 GN SC05447 OR SC3D11.04C.
 OS *Streptomyces coelicolor*.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Brown S.P., Harris D.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RC MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 RT the 8 Mb *Streptomyces coelicolor* A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howard S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomyete *Streptomyces*
 RT *coelicolor* A3(2)."
 RL Nature 417:141-147(2002).
 DR EMBL: AL157916; CAB76001.1; -
 DR HSSP: P05806; INPC.
 DR InterPro: IPR001570; Peptidase_M4.
 DR InterPro: IPR005075; Pep_M4_prop.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01447; Peptidase_M4; 1.
 DR Pfam: PF02868; Peptidase_M4; C. 1.
 DR Pfam: PF03413; Pep_M4_propep; 1.
 DR PRINTS: PR00730; THERMOLYSIN.
 DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
 DR Metalloprotease; Protease.
 SO SEQUENCE 684 AA; 71289 MW; 594F70DDAB732B8F CRC64;

Query Match 10.5%; Score 89; DB 16; Length 684;
 Best Local Similarity 27.6%; Pred. No. 5.2;
 Matches 40; Conservative 17; Mismatches 64; Indels 24; Gaps 4;

OY 37 SAYGV-----LSNFGTVKSGSSYPPTTSERPRV-VNNSRTDKPMV 80
 DB 453 SCIGGVVHYSSGPNHHPYLLSESGTITNGVYVNSPTSGCLVTVTGIGRAKAKIMFR 512
 OY 81 ALYLPVSSAGGLVTKAGSLIIVLLRQTNNTNSDDPQFVWNTYANNDDVYVPTGCD-- 137

```
Db 513 ALTTKFTSTNYAGARTGTLAAGELYGT---DSAEYTAANAMAGINVGARPGCGDPCD 569
QY 138 --VSARDVYTLDPYSGVPIPLTV 160
Db 570 GTTFESTINVINIPDYSAYVSSILTV 594

RESULT 20
Q9YN73 PRELIMINARY: PRT: 315 AA.
ID Q9YN73
AC Q9YN73
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
OS Hypothetical 34.8 kDa protein.
OC Vaccinia virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LIVP-2:
RA Babkin I.V., Petrov N.A., Shchelkunov S.N.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; 299045; CAB11743.1; -.
DR InterPro; IPR003599; 19.
DR InterPro; IPR003006; 19_MHC.
DR Pfam; PF00047; 19; 1.
DR SMART; SM00409; IG; 1.
DR Hypothetical protein.
SQ SEQUENCE 315 AA; 34836 MW; 97613E87D577708 CRC64;

Query Match
Best Local Similarity 22.7%; Score 88.5; DB 12; Length 315;
Matches 42; Conservative 21; Mismatches 49; Indels 73; Gaps 9;

QY 2 VVNVGONTLVVDLSTQIFCHN-----DY-----PEITIDVYTLQRGSAV 39
DB 127 IYNTDSESTIDILSGSTHSPETSEKRPYIDNSNCSSVFELATPEPTIDNV----- 178
QY 40 GGVLSNFGSTVYK-----SGSSYPPTTSETPRVVYNSRTDKPWPALYLPVSSAG 88
DB 179 ----EDHDTVYVTSDSINTVATSGES---TTDETPETPIDKEDHTVTDVTSYTVSS 230
QY 89 SAGGLVIRKAGSLIAVLILROTNNNNSDDFOFWNIYANDVYVPT--GGCDVSARDVYTL 146
DB 231 TSSGIV-----TTKSTTDAD-LYDYNNDNDIVPPTTVGGS-----TTS 268
QY 147 LPDYR 151
DB 269 ISNYK 273

RESULT 21
Q9YN68 PRELIMINARY: PRT: 270 AA.
ID Q9YN68
AC Q9YN68;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, last annotation update)
DE Hypothetical 29.8 kDa protein.
OS Cowpox virus (CPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HAMBURG-1985;
RA Babkin I.V., Petrov N.A., Shchelkunov S.N.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; 299050; CAB11748.1; -.
DR Hypothetical protein.
```

```
SQ SEQUENCE 270 AA; 29822 MW; 2B952EC06A8445E6 CRC64;
Query Match
Best Local Similarity 10.3%; Score 87; DB 12; Length 270;
Matches 41; Conservative 22; Mismatches 52; Indels 66; Gaps 9;

QY 2 VVNVGONTLVVDLSTQIFCHN-----DY-----PEITIDVYTLQRGSAV 39
DB 83 IYNTDSESTIDILSGSTHSPETSEKRPYIDNSNCSSVFELATPEPTIDNV----- 134
QY 40 GGVLSNFGSTVYKSGSSYPF-----PTTSETPRVVYNSRTDKPWPALYLPVSSAG 92
DB 135 ----EDHDTVYVTSDSINTVATSGEPTTDETPETI--TDKEDHTVTDVTSYTVSSG 189
QY 93 LYIKAGSLIAVLILROTNNNNSDDFOFWNIYANDVYVPT--GGCDVSARDVYTLDPY 150
DB 190 IV-----TTKSTTDAD-LYDYNNDNDIVPPTTVGGS-----TTSISNY 227
QY 151 R 151
DB 228 K 228

RESULT 22
Q9YN66 PRELIMINARY: PRT: 313 AA.
ID Q9YN66
AC Q9YN66;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
OS Hypothetical 34.4 kDa protein.
OC Monkeypox virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10244;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ZAIRE 77-0666;
RA Babkin I.V., Petrov N.A., Shchelkunov S.N.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; 299052; CAB11750.1; -.
DR InterPro; IPR003599; 19.
DR InterPro; IPR003006; 19_MHC.
DR Pfam; PF00047; 19; 1.
DR SMART; SM00409; IG; 1.
DR Hypothetical protein.
SQ SEQUENCE 313 AA; 34434 MW; 329993BF54B33CC CRC64;

Query Match
Best Local Similarity 10.1%; Score 85.5; DB 12; Length 313;
Matches 36; Conservative 17; Mismatches 37; Indels 51; Gaps 7;

QY 24 PEITIDVYTLQRGSAVSGVLSNFGSTVYK-----SGSSYPPTTSETPRVVYNS 72
DB 169 PEPTIDNV-----EDHDTVYVTSDSINTVATSGES---TTDETPETPIDK 212
QY 73 RTDKPWPALYLPVSSAGGLVIRKAGSLIAVLILROTNNNNSDDFOFWNIYANDVYVPT 132
DB 213 EEDHTVTDVTSYTVSSGIV-----TTKSTTDAD-LYDYNNDNDIVPPTTV 257
QY 133 T--GGCDVSARDVYTLDPDYR 151
DB 258 TTGGS-----TTSISNYK 271

RESULT 23
Q8V4S6 PRELIMINARY: PRT: 313 AA.
ID Q8V4S6
AC Q8V4S6;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE B2R.
```

```

GN B2R.
OS Monkeypox virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10244;
RN 11
RP SEQUENCE FROM N.A.
RC MEDLINE=21592287; PubMed=11734207;
RX Shchelkunov S.N., Tolmenin A.V., Babkin I.V., Saifonov P.F.,
RA Ryzankina O.I., Petrov N.A., Gutorov V.V., Uvarova E.A.,
RA Ryzankina O.I., Petrov N.A., Gutorov V.V., Uvarova E.A.,
RA Mkhoev M.V., Sisler J.R., Esposito J.J., Jahrling P.B., Moss B.,
RA Sandakchiev L.S.;
RT "Human monkeypox and smallpox viruses: genomic comparison.";
RL FEBS Lett. 509:66-70(2001).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=Zaire-96-I-16;
RA Shchelkunov S.N., Tolmenin A.V., Saifonov P.F., Gutorov V.V.,
RA Ryzankina O.I., Petrov N.A., Babkin I.V., Uvarova E.A.,
RA Esposito J.J., Moss B., Sisler J.R., Jahrling P.B., Sandakchiev L.S.;
RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF380138; AAL40621.1;
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00409; Ig_1.
SQ SEQUENCE 313 AA; 34452 MW; A7D916067F56ACF0 CRC64;

```

```

Query Match          10.1%; Score 85.5; DB 12; Length 313;
Best Local Similarity 25.5%; Pred. No. 4.2;
Matches 36; Conservative 17; Mismatches 37; Indels 51; Gaps 7;

```

```

OY 24 PETIDVYTLORGSAVGGVLSNFGSTVY-----SGSSYPTTSETPRVYNS 72
DB 169 PEPIDNV-----EDHTDTVTYSDSINTVASSGES---TTDETPEPIDK 212
OY 73 RDCDMPVALVLTTPVSSAGGLVYKAGSLIAVLILRQTNNSDDPQFVYANNDVVP 132
DB 213 EEDHTVTDVSTVSTSTSGIV-----TKSTTDDAD-LXDTINDNDVPP 257
OY 133 T--GGCDVSARDVYVTLDPYR 151
DB 258 TVVGS-----TTSISNYK 271

RESULT 24
O26812 PRELIMINARY; PRT; 1755 AA.
AC 026812;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Cell surface glycoprotein (S-layer protein).
GN MT4716.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).

```

```

DR EMBL; AE000851; AAB85221.1; -.
DR InterPro; IPR000601; PKD_domain.
DR Pfam; PF00801; PKD_2.
DR SMART; SM00089; PKD_2.
DR PROSITE; PS00093; PKD_2.
KW Complete proteome.
SQ SEQUENCE 1755 AA; 191961 MW; 02C446C99BB6F3DA CRC64;

```

```

Query Match          9.9%; Score 84; DB 17; Length 1755;
Best Local Similarity 23.9%; Pred. No. 51;
Matches 44; Conservative 25; Mismatches 61; Indels 54; Gaps 9;

```

```

OY 26 TITDVTYTLORGSAV-----GGVLSNFGSTVYKSSSY 58
DB 1423 TKENYITVYGPQTLENLVPTSGVAPLNTVSNLNTSGVAGNNTAQLKNCQTVA 1482
OY 59 FPTTS-----ETPRV-----VYNSRDKPMPVAL-YLTPVS-SAGGLVYKAGSLIV 103
DB 1483 TKTIAPACQTVSEFSRVLTAGVYNTIDLPVAVTVLKPANITVGNLTVPKSGVAP 1542
OY 104 LILRQTNVY-NSDDE-----QFVWN--IYANDVVPVPGGCVSARDVYVTLDPYRGSAV 154
DB 1543 LNTITVNLNRNTDLAGNTAQRKVGQVYAKTIAVPAGGVGVSPNTTLT---SAGNV 1599
OY 155 PIPV 158
DB 1600 PVAV 1603

```

```

RESULT 25
O05352 PRELIMINARY; PRT; 1114 AA.
AC 005352;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Immuno-dominant variable surface antigen.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=90354789; PubMed=1696956;
RA Edman U., Meraz M.A., Rausser S., Agabian N., Meza I.;
RT "Characterization of an immuno-dominant variable surface antigen from
RT pathogenic and nonpathogenic Entamoeba histolytica.";
RL J. Exp. Med. 173:879-888(1990).
DR EMBL; X55028; CAA38847.1; -.
DR InterPro; IPR000130; Zn_MTPetase.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 1114 AA; 125570 MW; 0445F99609313E1D CRC64;

```

```

Query Match          9.8%; Score 83.5; DB 5; Length 1114;
Best Local Similarity 23.6%; Pred. No. 32;
Matches 38; Conservative 32; Mismatches 58; Indels 33; Gaps 7;

```

```

OY 1 PVVWGNLVVDLSTQIFCHNDYPTITDVTYTLORGSAVGGVLSNFGSTVYKSSSY-PF 59
DB 214 PIDSGFDLGLN-TTQPIIND-----TFKIGSPFGMIYLRSDTT-FTNSPVFTF 262
OY 60 PTTSETPRVYVNSRDKMPVALYLTTPVSSAGGLVYKAGSLIAVLILRQTNVNSDDPQF 119
DB 263 SNVGKAPLIYNTTNEKNSVL-----RNAPGVVAELRTPGNLVLRSNIRSLDAQY 317
OY 120 VVNTIYANDVVPVPGGCVSARDVYVTLDPYRGSAVPIPLTV 160
DB 318 ISDFWLK-----AISISNVAVTLEN-----IPITL 342

```

```

RESULT 26
O09F47 PRELIMINARY; PRT; 315 AA.
ID O09F47
AC O09F47;

```

DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE TA66R.
 OS Vaccinia virus (strain Tian Tan).
 OC Viruses: dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 NCBI_TaxID=10253;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TIAN TAN;
 RA Jin Q., Hou Y.D., Cheng N.H., Yao E.M., Cheng S.X., Yang X.K.,
 RA Jing D.Y., Yu W.H., Yuan J.S., Ma X.U.;
 RT "Complete genomic sequence of vaccinia virus (Tian Tan strain).";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF095683; AAF34065.1;
 DR InterPro: IPR003599; 19;
 DR InterPro: IPR003006; 19_MHC.
 DR Pfam: PF00047; 19; 1.
 DR SMART: SM00409; 19; 1.
 SQ SEQUENCE 315 AA; 34758 MW; E04944814085CAA4 CRC64;

Query Match 9.7%; Score 82.5; DB 12; Length 315;
 Best Local Similarity 23.8%; Pred. No. 8.2;
 Matches 38; Conservative 22; Mismatches 43; Indels 57; Gaps 8;

QY 4 NWGONLVLDLSTQIFCHNDYPTITDVTYTLQGSAYGVLSNFGSTVKY----- 52
 DB 159 NINCSVFEIAT-----PEPIDNV-----EDHTDVTYTSISNITVSAS 198
 QY 53 GSSSYFPTTSETPRVYNSRDKPVPVALYLPVSSAGGLVTKAGSLIAYLILRQTNVY 112
 DB 199 SSES-----TTDETPEITDKEDHVTVDVSYTWTSTSGIV-----TTKS 240
 QY 113 NSDDFOFVNNIYANDVVPYPTG-GCDVSARDVTLLPDYR 151
 DB 241 TTDAD-LXDTYNDNDVFTVTCGS-----TTSISNKK 273

RESULT 27
 058048 PRELIMINARY; PRT; 1155 AA.
 AC 058048;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, last annotation update)
 DE 1155AA long hypothetical PYROLYSIN.
 GN PH0310.
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OT3;
 RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
 RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
 RA Masuchi Y., Shizuya H., Kikuchi H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
 RL DNA Res. 5:55-76(1998).
 DR EMBL: AP000001; BAA29383.1;
 DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000169; SHPOT_acsite.
 DR Pfam: PF00112; Peptidase_C1; 1.
 DR Pfam: PD000158; Peptidase_C1; 1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 1155 AA; 130098 MW; 092F4EC73144AF79 CRC64;

Query Match 9.7%; Score 82.5; DB 17; Length 1155;
 Best Local Similarity 24.5%; Pred. No. 42;
 Matches 46; Conservative 24; Mismatches 59; Indels 59; Gaps 11;
 QY 14 STQIFCHNDYPTITDVTYTLQGSAYGVLSNFGSTVYSGSSYFPTTSETPR----- 67
 DB 566 NTIVWFHN-YLDFWIDYIS---DKRYNALISLSTLESTLQAHFP---ESPAAFVDSQ 618
 QY 68 ---VYV---NSRTDKPVPVALYLP-----VSSAGGLVTKAGSLIAYL 104
 DB 619 LSDYLVYVMEKNSYPTETVEYQPKDILPNTGELYNTLLISIGGEYVLSL----- 674
 QY 105 ILRQTNVNSDDFOFV-----NNIYANDVVPYPTGCDVSARDVTLLPDYR--- 151
 DB 675 ---EGNVSIFPGGIVYVLEIPVVSVDV-APNNVTVQGNFNVSIVSI-IPLEDARIVI 728
 QY 152 GSVPIPLT 159
 DB 729 GNESYPLT 736

RESULT 28
 09X6L5 PRELIMINARY; PRT; 383 AA.
 ID 09X6L5;
 AC 09X6L5;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE Hypothetical 40.7 kDa protein.
 OS Klebsiella oxytoca.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.
 NCBI_TaxID=571;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=20334746; PubMed=10873523;
 RX Preston K.E., Radomski C.C.A., Venezia R.A.;
 RT "Nucleotide sequence of a 7-kb fragment of pACM1 encoding an IncM DNA
 RT plasmid and other putative proteins associated with conjugation.";
 RL Plasmid 44:12-23(2000).
 DR EMBL: AF139719; AAD33810.1;
 KW Hypothetical protein; Plasmid.
 SQ SEQUENCE 383 AA; 40703 MW; 5F338A503B65911A CRC64;

Query Match 9.7%; Score 82; DB 2; Length 383;
 Best Local Similarity 29.4%; Pred. No. 12;
 Matches 35; Conservative 18; Mismatches 54; Indels 12; Gaps 6;
 QY 47 SGTVKYS-GSSYP-----FPTTSETPRVYNSRDKPVPVA---LYLPVSSAGGLVTKAG 98
 DB 163 SQTVLSNAGSIFPVRFPNQATT--VTFSDANGHPVWLCAAPPYNSPSCASGGELCVGI 220
 QY 99 SLIAYLILRQTNVNSDDFOFVNNIYANDVVPYPTGCDVSARDVTLLPDYRGSVPIR 157
 DB 221 PGSAVFITOPTNAVASGNITVLLKGLA-TVYIINVKCAEBSVASKVDV-DYRLDLRIP 277

RESULT 29
 089120 PRELIMINARY; PRT; 315 AA.
 ID 089120;
 AC 089120;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE Hemagglutinin precursor.
 OS Vaccinia virus (strain Tian Tan).
 OC Viruses: dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 NCBI_TaxID=10253;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-TIAN TAN;
RA Li Z., Jin Q., Yu W., Jin T., Hou Y.;
RT "Complete nucleotide sequence of the hemagglutinin gene of tian tan
strain of vaccinia virus.";
RL Ping Tu Hsueh Pao 5:1-9(1989).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-TIAN TAN;
RA
XU L.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U25662; AAA74188.1; .
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00409; Ig; 1.
KW Signal.
FT SIGNAL
SQ SEQUENCE 315 AA; 34772 MW; 794DC90B57879D53 CRC64;

Query Match 9.6%; Score 81.5; DB 12; Length 315;
Best Local Similarity 23.8%; Pred. No. 10;
Matches 38; Conservative 22; Mismatches 43; Indels 57; Gaps 8;
QY 4 NVGQMLVVDLSTQIFCHNDYPTTIDYVTLQSGAVGVLNFGSTVY----- 52
DB 159 NNCSSVEIAT-----PEPIDNV-----EDHTDTVTYSDSIMTVSAT 198
QY 53 SCSSTPEPTTSETPRVYVNSRTDKPWPALVITPVSSAGGLVYKAGSLIALLIROTNNY 112
DB 199 SES-----TTDETPEPTIDKEEDHTVTDTVSYTVSTSGIV-----TTKS 240
QY 113 NSDDFOFVWNIYANNDVVPVPG-GCDVSAADVTATLPOVR 151
DB 241 TTDDGD-LYDIYNDNDVPTTVGCS-----TTSISNKK 273

RESULT 30
ID 082496 PRELIMINARY; PRT; 566 AA.
AC 082496;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DR 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hemagglutinin.
GN HA
OS Influenza virus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OX NCBI_TaxID=11320;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-A/SWINE/ANGE-GARDIEN/150/90(H3N2);
RX MEDLINE=95205091; PubMed=7897358;
RA Bikour M.H., Frost E.H., Deslandes S., Talbot B., Weber J.M.,
Elazhary Y.;
RT "Recent H3N2 swine influenza virus with haemagglutinin and
nucleoprotein genes similar to 1975 human strains.";
RL J. Gen. Virol. 76:697-703(1995).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: U07146; AAA85781.1; .
DR HSP: P03437; 2V1U.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTIN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63362 MW; 1FB8319A567E2EFF CRC64;

Query Match 9.6%; Score 81.5; DB 12; Length 566;

Best Local Similarity 26.4%; Pred. No. 21;
Matches 39; Conservative 17; Mismatches 57; Indels 35; Gaps 7;
QY 21 NDPEPTIDVYTLQSGAVGVL-----NFGSTVYKSSGSEPTTSETPRVVNSRT 74
DB 112 NCYPYDVPDYASLRSLSVAGSTLEFNEGFMTGVTONGSS---ACKRPDNGFPSRL 167
QY 75 DKPWPVALVITPVSSAGGLVYKAGSLIALLIROTNNVNSDPOFVNNY-----ANN 127
DB 168 N-W-----LY-----KSGSTYPQVQVTPPNNDSNKL-YINGVHHPTDKROT 209
QY 128 DVVPTGCGDVASADVTYTLDPYRGSPV 155
DB 210 LVQASGKVTSTKRSQGTVPNVGSR 237

RESULT 31
ID 040713 PRELIMINARY; PRT; 824 AA.
AC 040713;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DR 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Protein kinase.
GN OSBK10.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartidae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN (1)
RP SEQUENCE FROM N.A.
RA Zhao Y., Feng X.-H., Bottino P.J., Kung S.-D.;
RT "Molecular and biochemical characterization of a receptor
Serine/Threonine Kinase from Rice.";
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: L27821; AAA33915.1; .
DR InterPro: IPR001480; B_lectin.
DR InterPro: IPR000719; Euk-kinase.
DR InterPro: IPR003014; PAN.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR004040; STY_pkinase.
DR Pfam: PF01453; Agglutinin; 1.
DR Pfam: PF00024; PAN; 1.
DR Pfam: PF00063; Pkinase; 1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00108; B_lectin; 1.
DR SMART: SM00221; STYK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 824 AA; 89870 MW; EEBE8440BEBDD8 CRC64;

Query Match 9.6%; Score 81; DB 10; Length 824;
Best Local Similarity 23.4%; Pred. No. 38;
Matches 39; Conservative 29; Mismatches 57; Indels 42; Gaps 9;
QY 7 ONLVV-DLSTQIFCHNDYPTTIDYVTLQSGAV--GVLNFGSTVY----- 51
DB 145 QSVLVGDKASSPLMQSFHP--TD--TLLSGQNFTEGMFLMSKNTVQNNYTLQIKSG 199
QY 52 ----YSGSSYPP--TSETPRVYVNSRTDKPWPALVITPVSSAGGLVYKAGSLIALLI 105
DB 200 NMILVAGFETQPVYSAOODSRITVYKNGDRITPANL-----SSWSWSPYDQSGILLQV 255
QY 106 LROTNN-----YNSDDFOFVWNIYANN-----DVVPTGCGDVSA 140
DB 256 IAGEANNTLSAVLGSGLIAFYMLQGGNGSKSFITVPADSCMPPA 302

RESULT 32

09SE00
ID 09SE00 PRELIMINARY: PRT: 473 AA.
AC 09SE00;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Purple acid phosphatase precursor (EC 3.1.3.2).
GN PAP1.
OS Ipomoea batatas (Sweet potato) (Batales).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.
OX NCBI_TaxID=4120;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99441212; PubMed=10510276;
RA Schenk G., Ge Y., Carrington L.E., Wynne C.J., Searle I.R.,
RA Carroll B.J., Hamilton S., de Jersey J.,
RT "Nuclear metal centers in plant purple acid phosphatases: Fe-Mn in
RT sweet potato and Fe-Zn in soybean."
RL Arch. Biochem. Biophys. 370:183-189(1999).
CC -I- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O = AN
CC ALCOHOL + PHOSPHATE.
DR EMBL: AF200825; AAF19821.1; -.
DR HSP; P80366; 4KBP.
DR InterPro: IPR004843; M-peptidase.
DR InterPro: IPR004844; S/T_phosphatase.
DR Pfam: PF00149; Metallophos. 1.
KW Hydrolase; Iron; Signal; Zinc.
FT SIGNAL 1 38
FT CHAIN 39 473
SQ SEQUENCE 473 AA; 53815 MW; BAE4B807DAD95A7 CRC64;

Query Match 9.5%; Score 80.5; DB 10; Length 473;
Best Local Similarity 37.9%; Pred. No. 21;
Matches 22; Conservative 12; Mismatches 19; Indels 5; Gaps 2;

QY 30 YVTTLQGSAYGGVLSNFGSTVYKSGSSYPPTSETPRVYNSRTDKPWPVALYLPV 87
DB 274 WVAIKRASHIIVLSSYSGFVKYPOYKWF--TSELEKY---NRSETPLIYLVNAPL 326

RESULT 33

09XC46 PRELIMINARY: PRT: 1304 AA.

AC 09XC46;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Outer membrane protein B (Fragment).
GN OMPB.
OS Rickettsia australis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PHS;
RX MEDLINE=20487299; PubMed=11034486;
RA Stenos J., Walker D.H.,
RT "The rickettsial outer-membrane protein A and B genes of Rickettsia
RT australis, the most divergent rickettsia of the spotted fever group."
RL Int. J. Syst. Evol. Microbiol. 50:1775-1779(2000).
DR EMBL: AF149109; AAD39532.1; -.
DR InterPro: IPR003858; romPA_ompB.
DR InterPro: IPR001639; T2SP_C.
DR Pfam: PF02708; romPA_ompB. 1.
DR PRINTS: PR00810; BCTERIALGSPC.
DR PRINTS: PR01656; VACCYOTOXIN.
FT NON_TER 1
SQ SEQUENCE 1304 AA; 133283 MW; ABFF9AAEED8A7701 CRC64;

Query Match 9.5%; Score 80.5; DB 2; Length 1304;

Best Local Similarity 25.4%; Pred. No. 75;
Matches 33; Conservative 13; Mismatches 53; Indels 31; Gaps 4;
QY 27 ITDVTTLQGSAYGGVLSNFGSTVYKSGSSYPPTSETPRVYNSRTDKPWPVALYLPV 86
DB 595 IVNATTLVAGI---GTTKNNGCTVTLSCG-----IPNTPGTIYGLGVEHGYPKIKQVTF 645
QY 87 VSSAGGLVTKAGSLIAVLILKOTNNYNSDDPQFVWNITANDVVPVTCGCVSARDVTVT 146
DB 646 ITDYNL-----GSIAT-----NVIINDGVYVTTGCVGTDFDGTIT 683
QY 147 LPDYRGSVPI 156
DB 684 LGSVNGNANV 693

RESULT 34

09KKB5 PRELIMINARY: PRT: 1620 AA.

AC 09KKB5;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE OMPB (Fragment).
GN OMPB.
OS Rickettsia australis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PHILIPS;
RX MEDLINE=20393643; PubMed=10939649;
RA Roux V., Raoult D.,
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RT gene coding the outer-membrane protein romPB (ompB)."
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
DR EMBL: AF123709; AAF34112.1; -.
DR InterPro: IPR003858; romPA_ompB.
DR Pfam: PF02708; romPA_ompB. 1.
FT NON_TER 1620 1620
SQ SEQUENCE 1620 AA; 164420 MW; F9E9DB056A548A1C CRC64;

Query Match 9.5%; Score 80.5; DB 2; Length 1620;
Best Local Similarity 25.4%; Pred. No. 99;
Matches 33; Conservative 13; Mismatches 53; Indels 31; Gaps 4;

QY 27 ITDVTTLQGSAYGGVLSNFGSTVYKSGSSYPPTSETPRVYNSRTDKPWPVALYLPV 86
DB 918 IVNATTLVAGI---GTTKNNGCTVTLSCG-----IPNTPGTIYGLGVEHGYPKIKQVTF 968
QY 87 VSSAGGLVTKAGSLIAVLILKOTNNYNSDDPQFVWNITANDVVPVTCGCVSARDVTVT 146
DB 969 ITDYNL-----GSIAT-----NVIINDGVYVTTGCVGTDFDGTIT 1006
QY 147 LPDYRGSVPI 156
DB 1007 LGSVNGNANV 1016

RESULT 35

09YN72 PRELIMINARY: PRT: 270 AA.

AC 09YN72;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
DE Hypothetical 29.8 kDa protein.
OS Vaccinia virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10245;
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-LIVP-1;
 RA Babkin I.V., Petrov N.A., Shchelkunov S.N.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: 299046; CAB11744.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 270 AA; 29798 MW; EADCOFIB0567CC2 CRC64;

Query Match 9.4%; Score 80; DB 12; Length 270;
 Best Local Similarity 22.7%; Pred. No. 12;
 Matches 42; Conservative 21; Mismatches 48; Indels 74; Gaps 10;

OY 2 VVNVGNLVNLSIQIFCHN-----DY-----PETITDVTLQSGAV 39
 DB 83 IVNTDSESTIDILSGTHSPETSEKPDYIDNSCNSVFEIATPEPTDNV----- 134
 OY 40 GGVLSNFGTVKY-----SGSSYPPTTSETPRVVYNSRTDKPMPVALYLPVS 88
 DB 135 ----EDHTDVTYTSDSINTVSASGES-----TTDEPEPI-TDKEDHTVTDVTSYTVS 185
 OY 89 SAGGLVITAGSLIIVLILRQTNNTNSDDEQFVWNIYANDVYVPT--GGCDVSARDVTVT 146
 DB 186 TSSGIY-----TTKSTTDDAD-LYDTYNDNDVTPPTVGS-----TTS 223
 OY 147 LPDYR 151
 DB 224 ISNYK 228

RESULT 36
 096UK8 PRELIMINARY; PRT; 519 AA.

ID 096UK8;
 AC 096UK8;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Laccase 1 (EC 1.10.3.2).
 GN LAC1.
 OS Trametes versicolor (White-rot fungus).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Aphyllophorales; Trametes.
 OX NCBI_TaxID=5325;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CUL;
 RA McGeehan K.H., O'Brien M.M., Dobson A.D.W.;
 RL "Trametes versicolor laccase (lac1) mRNA sequence."
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY049725; AAL00887.1; -
 DR InterPro: IPR001117; Cu-oxidase.
 DR Pfam: PF00394; Cu-oxidase; 3.
 DR PROSITE: PS00079; MULTICOPEER_OXIDASE1; UNKNOWN_1.
 KW Oxidoreductase.
 SQ SEQUENCE 519 AA; 55998 MW; 8C9C6934DEAE3E0A CRC64;

Query Match 9.4%; Score 80; DB 3; Length 519;
 Best Local Similarity 27.7%; Pred. No. 27;
 Matches 38; Conservative 21; Mismatches 48; Indels 30; Gaps 8;

OY 7 QNLVLDLSIQIFCHNDY-----ETITDVTLQSGAVGGVLSNFGTV-----KYSGS 55
 DB 249 QPLVLD-SIQIFAAQRISFVLNAGVGNV--WVRANPFGTV--GFGAGINSAILRYQGA 304
 OY 56 STEPPTTSETPRVVYNSRTDKPMPVALYLPVSAGGLVITAGSLIIVLILRQTNNTNSD 115
 DB 305 PVAEPPTTQTPSVIPIETNL-HPLAMVPGRTRPGVDKA-----LKLAEFNGT 355
 OY 116 DFOFVWNIYANDVYVPT 132
 DB 356 NF-----FINNASFTP 366

RESULT 37

089182
 ID 089182 PRELIMINARY; PRT; 315 AA.

AC 089182;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HAEMAGGLUTININ.
 GN A56R OR MVA165R.
 OS Vaccinia virus (strain Ankara).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=126794;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97080498; PubMed=8921843;
 RA Antoline G., Scheifflinger F., Holzer G., Langmann T., Falkner F.G.,
 RA Donner F.;
 RT "Characterization of the vaccinia MVA hemagglutinin gene locus and its
 RT evaluation as an insertion site for foreign genes."
 RL Gene 177:43-46(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ANKARA;
 RA Antoline G., Scheifflinger F., Falkner F.G., Donner F.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X91135; CA62575.1; -
 DR EMBL: U94848; AAB96543.1; -
 DR InterPro: IPR003599; Ig_MHC.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig; 1.
 DR SMART: SM00409; IG; 1.
 SQ SEQUENCE 315 AA; 34794 MW; C7EE7D42D7E4E543 CRC64;

Query Match 9.4%; Score 79.5; DB 12; Length 315;
 Best Local Similarity 22.2%; Pred. No. 16;
 Matches 41; Conservative 21; Mismatches 50; Indels 73; Gaps 9;

OY 2 VVNVGNLVNLSIQIFCHN-----DY-----PETITDVTLQSGAV 39
 DB 127 IVNTDSESTIDILSGTHSPETSEKPDYIDNSCNSVFEIATPEPTDNV----- 178
 OY 40 GGVLSNFGTVKY-----SGSSYPPTTSETPRVVYNSRTDKPMPVALYLPVS 88
 DB 179 ----EDHTDVTYTSDSINTVSASGES-----TTDEPEPI-TDKEDHTVTDVTSYTVS 230
 OY 89 SAGGLVITAGSLIIVLILRQTNNTNSDDEQFVWNIYANDVYVPT--GGCDVSARDVTVT 146
 DB 231 TSSGIY-----TTKSTTDDAD-LYDTYNDNDVTPPTVGS-----TTS 268
 OY 147 LPDYR 151
 DB 269 ISNYK 273

RESULT 38
 067126 PRELIMINARY; PRT; 566 AA.

ID 067126;
 AC 067126;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hemagglutinin.
 GN HA.
 OS Influenza A.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses.
 OX NCBI_TaxID=11320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A/SEAL/MA/3984/92;
 RX MEDLINE=95146951; PubMed=7844533;
 RA Callan R.J., Early G., Kida H., Henshaw V.S.;
 RT "The appearance of H3 influenza viruses in seals."

```

RL J. Gen. Virol. 76:199-203(1995).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: I32024; AAA64228.1; -.
DR HSSP: P03437; 2VI0.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin: 1.
DR PRINTS: PR00329; HEMAGGLUTIN12.
DR ProDom: PD000225; Hemagglutin; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63441 MW; 590576CB4CE7D08 CRC64;

Query Match
Best Local Similarity 9.4%; Score 79.5; DB 12; Length 566;
Matches 38; Conservative 17; Mismatches 58; Indels 35; Gaps 7;

QY 21 NDYPETITDVTVLQSGSAYGVLS-----NFGTVKSGSSYPPTTSETPRVVYNSRT 74
DB 112 NCPYDVPYASLSRLVYASSGTLEFIAGEFTWGTQNGSG-----TCKRGFANGFFSL 167
QY 75 DKRPVALTLTPYSSAGGLVYKAGSLIAVLILRQTNVNSDDEFOFYWNIT---ANND--- 128
DB 168 N-W-----LTRKSGAYPVLNTMPNNDNDKL-YWGVHHSSTQEQTS 209
QY 129 -YVPTGCDVSARDVTTLTPYRGSVP 155
DB 210 LVQASGRVTVSTRSQOTIINIGSSP 237

RESULT 39
Q9JFS2 PRELIMINARY; PRT; 1924 AA.
AC Q9JFS2
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15; Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19; Last annotation update)
DE C15R.
GN C15R.
OS Ectromelia virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=12643;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MOSCOM.
RX MEDLINE=20192152; PubMed=10725549;
RA Chen N., Buller R.M.L., Wall E.M., Upton C.;
RA "Analysis of host response modifier ORFs of ectromelia virus, the
RA causative agent of mousepox.";
RL Virus Res. 66:155-173(2000).
DR EMBL: AF012825; AAC9574.1; -.
SQ SEQUENCE 1924 AA; 216775 MW; B34C50FEA260BB95 CRC64;

Query Match
Best Local Similarity 9.4%; Score 79.5; DB 12; Length 1924;
Matches 46; Conservative 36; Mismatches 73; Indels 99; Gaps 10;

QY 1 PVVNYGQNLVYDLSTQ-----IFCHNDPE-----TITD----- 29
DB 1286 PVGDGIATKLVSATKQSIKAVFCHNDRKRSIDTILFIENSIDIRSSITGDNCRN 1345
QY 30 --YVTLQSGAYGV-LSNFSGTVYSGSSY-----PFPTTSEPR 67
DB 1346 RLFTLSGMYRSMYCNRRPGCTYAGIPFNENSVESDLHGPETIMUKETIDANSIDPR 1405
QY 68 VVNSRDTKMPVVALYLPVSSAGS-----LVTKA----- 97
DB 1406 VITKSKTHFPPIISVKEFVAVNLGNGYKNPEFVNDKSKKRTYSAMTIKILPCTVRNKNV 1465
QY 98 -----GSLIAVLILRQTNVNSDDEFOFYWNITVANDVYVPTGCC-----DVSARDVTVL 147
```

```

DB 1466 NFGYNGHILSNMVAQSTSYDGD-----GTWYTKSVKRSDECSJLDLAKETVTC 1521
QY 148 PDYRGSVPITPTVY 161
DB 1522 PAF--SIPRISAY 1533

RESULT 40
Q64302 PRELIMINARY; PRT; 421 AA.
AC Q64302
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07; Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08; Last annotation update)
DE A-protein.
OS bacteriophage M11.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Leviviridae;
OX Allollevivirus.
OX NCBI_TaxID=74336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95239761; PubMed=7723040;
RA Beekwilder M.J., Nieuwenhuizen R., van Duin J.;
RT "Secondary structure model for the last two domains of single-stranded
RT RNA phage Q beta.";
RL J. Mol. Biol. 247:903-917(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96190948; PubMed=8609616;
RA Beekwilder J., Nieuwenhuizen R., Poot R., van Duin J.;
RT "Secondary structure model for the first three domains of Q beta RNA.
RT J. Mol. Biol. 256:8-19(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Beekwilder J., Nieuwenhuizen R., Poot R., van Duin J.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF052431; AAC06249.1; -.
SQ SEQUENCE 421 AA; 48619 MW; BF24627BF4EC7707 CRC64;

Query Match
Best Local Similarity 9.3%; Score 79; DB 9; Length 421;
Matches 34; Conservative 15; Mismatches 57; Indels 24; Gaps 4;

QY 31 VTIQGSAYGVLSNFSGTAVY-SGSSYPTTSETPRVYNSRTDKPVPVALYLPVSS 89
DB 267 ITVORRHRWGLIVANREGYATFDNGSIRPVSDMKELANAFIN-----PGEVAMELTPYSR 321
QY 90 AGGIVIKAGSLIAVLILRQTNVNSDDEFOFYWNITVANDVYVPTGCDVSARDVTLPD 149
DB 322 VVDFMFINVDIIL-----EQQKQLXQNDIVDYGQRDIRNR--SVTLKG 363
QY 150 YRGSVPITPT 159
DB 364 VRNGIPVAVT 373

RESULT 41
Q9VFW7 PRELIMINARY; PRT; 985 AA.
AC Q9VFW7
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13; Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21; Last annotation update)
DE CG8775 protein.
GN CG8775.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
```


RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RA MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amenatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Berens P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
 RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Chertys J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dev I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
 RA Foster A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jajall M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kamel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinett K., Remington K., Saunders R.D.C., Scheeler F., Shen T.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith H.,
 RA Sjöberg R., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrstka R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003699; AFS4928.1;
 DR FlyBase: FBgn0038138; CG8775.
 DR InterPro: IPR001930; Ala-peptase.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01433; peptidase_M1; 1.
 DR PRINTS: PR00756; ALADIPASE.
 DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
 SQ SEQUENCE 985 AA; 11664 MW; 56CAC0CF41172224 CRC64;

Query Match 9.38; Score 79; DB 5; Length 985;
 Best Local Similarity 21.98; Pred. No. 73;
 Matches 43; Conservative 24; Mismatches 57; Indels 72; Gaps 9;

QY 1 PVVNVGNLVVDLSTQIFCHNDYPTITDY---VTLRGSGYGVGLNFSSTVYKSSSY 57
 DB 502 PVLITDITFLA---SHPIYKSIESPALTEYFDITITSGALVLMLENGEELRNA-- 556
 QY 58 PEPTSETPRVYNSRTDKPWPVALYLPVSSAGL-----VIK 96
 DB 557 ---TTRYLVRHIYSTATEDD---YLVAEEGLEFDVQIMQMTWTEOKGLPVEVER 608
 QY 97 AGS---LIVLILQOTNNY-----NSDQFQPVWN-----IYANDVVVPT 133
 DB 609 SGGSTYKLTQRKFLANEDVLAEAASFNRYMSPITTYTSINSEVOSLIENHND----- 663
 QY 134 GGCDSARDVYTLPD 149
 DB 664 -----NEATITLPE 672

RESULT 42
 Q9FXI9

ID Q9FXI9 PRELIMINARY; PRT: 1064 AA.
 AC Q9FXI9;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative protein kinase.
 GN F12G12.7.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 [1]
 RP SEQUENCE FROM N.A.
 RA Federle H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
 RA Alatali H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
 RA Buehler E., Chao Q., Chin C., Chou J., Choi E., Gonzalez A.,
 RA Howng B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
 RA Lenz C., Liu A., Liu S., Mukharbeky N., Pham P., Sakano H., Shinn P.,
 RA Tortum M., Vayberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC015446; AAC12526.1;
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR.
 DR InterPro: IPR004838; NHRtransf_1.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR004040; STY_pkinase.
 DR InterPro: IPR001245; TYR_pkinase.
 DR Pfam: PF00560; LRR; 24.
 DR Pfam: PF00069; pkinase; 1.
 DR PRINTS: PR00019; LEURICHRPT.
 DR Prodom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00370; LRR; 23.
 DR SMART: SM00221; STYK; 1.
 DR SMART: SM00220; STYK; 1.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00105; AA_TRANSFER_CLASS_1; UNKNOWN_1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_STY; 1.
 KW ATP-binding; Kinase; Transferase.
 SQ SEQUENCE 1064 AA; 115915 MW; 2E8FC24DEF4F053D CRC64;

Query Match 9.38; Score 79; DB 10; Length 1064;
 Best Local Similarity 21.28; Pred. No. 81;
 Matches 45; Conservative 21; Mismatches 58; Indels 88; Gaps 10;

QY 11 VDLSTQIFCHNDYPTITDYVTLQ-----RGSAYGV-----LSNFGTV 50
 DB 602 LDIYSNFTGN-IPETPSDLTQQLQSLDLSNSLHGDIKVLGSLTSLASLNSCNHNSGPI 660
 QY 51 KYSSSYPEPTSTSTPRVYNSR-----TDKPPVALYLPVSSAG 91
 DB 661 ---PSTPEFTTSTSYLQNTNLCHSLDITCSSHTGONGVKSPTIVAL----- 707
 QY 92 GLVIKAGSLIAV-----LIDRFNNY-----NSDQFQPVNIYANDVVVPTG 135
 DB 708 TAVILASITAILAIAWLLILRNHLTYTSQNSSSPSTADDFSPWTF-----IPFK 760
 QY 136 CDVSARDVYTLPD-----YRGSVP 155
 DB 761 LGITVNNIVTSLDENYIGKSGIYKAEIP 792

RESULT 43
 Q8T2A1
 ID Q8T2A1 PRELIMINARY; PRT: 3763 AA.
 AC Q8T2A1;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 416.9 kDa protein.

```

OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
ON NCBI_TaxID=44689;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
  Lehmann R., Baumgart C., Parra G., April J.F., Gulgo R., Kumpf K.,
  Tunngal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
  "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
  Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RT EMBL: AC115678; AAL92607.1; -
DR Hypothetical protein.
SQ SEQUENCE 3763 AA; 416939 MW; 83E8D27C0164EDF6 CRC64;

Query Match
Best Local Similarity 24.5%; Pred. No. 3.9e+02;
Matches 46; Conservative 25; Mismatches 75; Indels 42; Gaps 8;

OY 5 VGQNLVDSLQIFCHN-----DYPTITDYVTLQRG-SAYGVLNFSGVTKISGSS-- 56
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1853 IGDTSVDLSNVIAGGSTVDGVDIPNVYIAGSGTVNGTGTSTTTVSSSPSSSS 1912
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 57 -----YFPTTSEPRVYNSRTDKP-----WPVALYLPV--SSAG 91
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1913 DIGSSDISSEVSSLSPPSSSEOPSEOSSSEOPEDSMEETPVH-YFEPVETTPG 1971
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 92 GLVIKAGSLIAVLILRQTNVNSDDFQFVWNIYANNVYPTGCDVSARDVTLTPDYR 151
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1972 SYPLDSIAERDILRNALVYARAPYAFGQKVF-----FIDNNGNEGVDPVTVLVNSN 2026
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 152 GSVPIPLT 159
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2027 G---FPVT 2031
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 44
O80901 PRELIMINARY; PRT; 398 AA.
AC O80901;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Penicillin-binding protein, dact.
DA DACT OR ATU3634 OR AGR_L_2387.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
ON [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
  Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
  Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D. St.,
  Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
  Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
  Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
  Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
  Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
  Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
  Nester E.W.;
  "The genome of the natural genetic engineer Agrobacterium tumefaciens
  C58.";
  Science 294:2317-2323 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
  Qurollo K., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
  Hummel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu F.,
  Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Marckel B.,
  Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,

```

```

RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
  Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328 (2001).
DR EMBL: AE009292; AAL44446.1; ALT_INIT.
DR EMBL: AE008320; AAK89763.1; -
KW Complete proteome.
SQ SEQUENCE 398 AA; 42160 MW; 8155CABE9565C5D CRC64;

Query Match
Best Local Similarity 24.2%; Pred. No. 26;
Matches 43; Conservative 26; Mismatches 64; Indels 45; Gaps 8;

OY 1 PVAVGQNLVYDLSLT-QIFCHN-----YPTITDYVTLQGSAYGVLNFSGVTKYSG 54
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 56 PVANANPRMVDVKTGIHQEAFKRYKYPASLTIKMT-----AYIAFSOMKAKTI----- 106
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 55 SSYPPTTSEPRVYNSRTDKPMPVALY-----LTPYSSAGLVIRAGSLIAVLILRQT 109
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 107 -----SPQTEVMSKKAADQPAKMYFKPGOKLTMDSALKLLIKSANDIAVAI-AET 158
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 110 NNYSDDFQFVWNIYAN-----NDVYVPTGCDVSARDVTV-----TLPDY 150
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 159 IGTSDNFAQMNQAQRIGMSSTHYVNPGLPGQGYTTARDLALLIKREFPEX 216
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 45
O93SH4 PRELIMINARY; PRT; 782 AA.
AC O93SH4;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hmur protein.
GN HMUR.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
ON [1]
RP SEQUENCE FROM N.A.
RC STRAIN=110SPC4;
RA Nienaber A., Hennecke H., Fischer H.M.;
  "Discovery of a haem uptake system in the soil bacterium
  Bradyrhizobium japonicum.";
  Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AJ311165; CAC38746.1; -
DR InterPro: IPR000531; TONB_BoxC.
DR Pfam: PF00593; TONB_BoxC; 1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
SQ SEQUENCE 782 AA; 83713 MW; F8FC1EBFE568AA87 CRC64;

Query Match
Best Local Similarity 25.0%; Pred. No. 61;
Matches 37; Conservative 17; Mismatches 65; Indels 29; Gaps 7;

OY 21 NDYPTITDYVTLQGSAYGVLNFSGVTKYSSSPPTTSEPRVYNSRTDKPWP 79
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 313 NGCPTTQALLILNKGSSVYASDAKNGYGVWV-YSLPSDLFPWHSYVGNGRFDNDQT 371
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 80 VALY--LTP-----VSSAGLVIRAGSLIAVLILRQTNNSDDFQ--FV 120
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 372 KYHYGTTPSAYCAVNGCFGNVSGVCDKRGVILNTRYGDA---NNTTFENVGDMRNALT 427
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 121 WNIYANNVYPTGCDVSARDVTLTP 148
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 428 WGVDAFPQDDVITTD-----SRGNSNITTP 451
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Search completed: November 28, 2002, 17:21:30
Job time : 30 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 17:17:25 : Search time 44 seconds
(without alignments)
351.765 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186

Perfect score: 848
Sequence: 1 PVVNVGONLVVDLSTQIFCH.....DVTVTLPDYKSGVPIPLFYV 161

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database :

1: PIR-73:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	834	98.3	300	2 S56545	fimbrial protein f
2	824	97.2	300	2 G91288	hypothetical prote
3	824	97.2	300	2 B86130	hypothetical prote
4	756	89.2	302	2 A32801	fimbrial adhesin p
5	384	45.3	304	2 A64904	probable fimbrial
6	384	45.3	304	2 C90892	probable adhesin
7	384	45.3	304	2 F85725	probable adhesin
8	207	24.4	299	2 C49233	S fimbrial adhesin
9	205	24.2	291	2 S15927	stah protein precu
10	199.5	23.5	299	2 I76900	Fl652 minor fimbri
11	86.5	10.2	550	1 HMIY98	hemagglutinin prec
12	85.5	10.1	315	1 HNVZVT	hemagglutinin prec
13	85.5	10.1	550	1 HMIY33	hemagglutinin prec
14	84	9.9	1797	2 F69195	cell surface glyco
15	83.5	9.8	367	1 HMIYV	hemagglutinin prec
16	83.5	9.8	639	2 A32935	protein PI - Entam
17	83.5	9.8	1114	2 JH0284	125K surface antiq
18	82.5	9.7	315	1 HNVZVT	hemagglutinin prec
19	82.5	9.7	1155	1 H71456	probable pyrolysin
20	81.5	9.6	550	1 HMIY33	hemagglutinin prec
21	81	9.6	824	1 S50767	S-reeceptor kinase
22	80.5	9.5	473	1 A59200	acid phosphatase
23	80	9.4	314	1 JQ1793	hemagglutinin prec
24	80	9.4	314	1 HNVZVT	hemagglutinin prec
25	80	9.4	385	2 S55524	neurokinin 3 recep
26	79.5	9.3	373	2 T37438	hemagglutinin - va
27	79	9.3	373	2 JH0155	pectin lyase (EC 4
28	79	9.3	1064	2 B86465	probable protein k
29	78.5	9.3	389	2 AH3003	penicillin-binding

ALIGNMENTS

RESULT 1

S56545
fimbrial protein fimb precursor, type 1 - Escherichia coli (strain K-12)

C:Species: Escherichia coli
C:Date: 28-Oct-1995 #sequence: revision 03-Nov-1995 #text_change 01-Mar-2002
C:Accession: S56545; B65246; S09563; A36967
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995

A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from
A:Reference number: S56314; MUID:95334362; PMID:7610040

A:Accession: S56545
A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-300 <BUR>

A:Cross-references: EMBL:U14003; NID:91263172; PIDN:AAA97216.1; PID:9537161
A:Note: the nucleotide sequence was submitted to the EMBL data library, August 1994
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426517; PMID:9278503

A:Accession: B65246
A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-300 <BUR>

A:Cross-references: GB:AE000502; GB:U00096; NID:92367374; PIDN:AACT7276.1; PID:917907
A:Experimental source: strain K-12, substrain MG1655
R:Klemm, P.; Christiansen, G.
Mol. Gen. Genet. 208, 439-445, 1987

A:Title: Three fimb genes required for the regulation of length and mediation of adhes
A:Reference number: S07321; MUID:88038337; PMID:2850081

A:Accession: S09563
A:Molecule type: DNA
A:Residues: 1-196; 'R', 198-221, 'H', 223-300 <RLE>

A:Cross-references: EMBL:X05672; NID:941463; PIDN:CAA29156.1; PID:941466
R:Sokurenko, E.V.; Courtney, H.S.; Ohman, D.E.; Klemm, P.; Hasty, D.L.
J. Bacteriol. 176, 748-755, 1994

A:Title: Fimb family of type 1 fimbrial adhesins: functional heterogeneity due to min
A:Reference number: A36967; MUID:94131954; PMID:7905476

A:Accession: A36967
A:Status: not compared with conceptual translation

A:Molecule type: nucleic acid
A:Residues: 1-78, 'R', 80-221, 'H', 223-300 <SOS>

A:Note: Sequence extracted from NCBI backbone (NCBIP:143314)

C:Genetics:
A:Gene: fimb
C:Function:
A:Description: involved in longitudinal regulation and mannose-specific adhesion
A:Note: controls length and number of fimbriae
C:Superfamily: fimbrial protein fimb

C:Keywords: fimbria

A:Residues: 1-639 <TAN>

Query Match 9.8%; Score 83.5; DB 2; Length 639;

Best Local Similarity 23.6%; Pred. No. 8.1;

Matches 38; Conservative 32; Mismatches 58; Indels 33; Gaps 7;

QY 1 PVNVVGMVVDLSTQIFCHNDYPTITDYTLQRGSAVGVLSNFGSTVYKSGSSY-PF 59

Db 28 PIDSIGFDLGIN-TTQPIIIND-----TFKIGSPFGGMITLRSDPT-FTNSFYVTF 76

QY 60 PTSTPTPVVYNSRDKPWPALYLTTPVSSAGLVKAGSLIAVLILRQTNVNSDDFOF 119

Db 77 SNVGRAPLIANNITTNEMNSVL-----RNAFGVAEIRTPGNRLVLSRNIKRSLEDAQY 131

QY 120 VNNIYANNDDVVPPTGGCDVSARDVYTLIPDRGSAVPIPLTV 160

Db 132 ISDFWLK-----AISISNVAVTLEN-----IPITL 156

RESULT 17

JH0284

125K surface antigen M17 precursor - Entamoeba histolytica

C:Species: Entamoeba histolytica

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999

C:Accession: JH0284

R:Edman, U.; Meraz, M.A.; Rausser, S.; Agabian, N.; Meza, I.

J. Exp. Med. 172, 879-888, 1990

A:Title: Characterization of an immuno-dominant variable surface antigen from pathogenic

A:Reference number: JH0284; MUID:90354789; PMID:1696956

A:Accession: JH0284

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1114 <EDM>

A:Experimental source: strain HM1 IMSS

Query Match 9.8%; Score 83.5; DB 2; Length 1114;

Best Local Similarity 23.6%; Pred. No. 16;

Matches 38; Conservative 32; Mismatches 58; Indels 33; Gaps 7;

QY 1 PVNVVGMVVDLSTQIFCHNDYPTITDYTLQRGSAVGVLSNFGSTVYKSGSSY-PF 59

Db 214 PIDSIGFDLGIN-TTQPIIIND-----TFKIGSPFGGMITLRSDPT-FTNSFYVTF 262

QY 60 PTSTPTPVVYNSRDKPWPALYLTTPVSSAGLVKAGSLIAVLILRQTNVNSDDFOF 119

Db 263 SNVGRAPLIANNITTNEMNSVL-----RNAFGVAEIRTPGNRLVLSRNIKRSLEDAQY 317

QY 120 VNNIYANNDDVVPPTGGCDVSARDVYTLIPDRGSAVPIPLTV 160

Db 318 ISDFWLK-----AISISNVAVTLEN-----IPITL 342

RESULT 18

HN2VYT

hemagglutinin precursor - vaccinia virus (strain Tiancan)

C:Species: vaccinia virus

C:Date: 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change 16-Jul-1999

C:Accession: JI0108

R:Jin, D.; Li, Z.; Jin, Q.; Yuwen, H.; Hou, Y.

J. Exp. Med. 170, 571-576, 1989

A:Title: Vaccinia virus hemagglutinin. A novel member of the immunoglobulin superfamily.

A:Reference number: JI0108; MUID:89328331; PMID:2754392

A:Accession: JI0108

A:Molecule type: mRNA

A:Residues: 1-315 <JIN>

A:Cross-references: GB:X15709; GB:M57773; NID:661313; PIDN:CAA33740.1; PID:961314

C:Superfamily: vaccinia virus hemagglutinin; immunoglobulin homology

C:Keywords: glycoprotein; hemagglutinin; late protein; transmembrane protein

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-315/Product: hemagglutinin #status predicted <MAT>

F:27-105/Domain: immunoglobulin homology <IMM>

F:280-302/Domain: transmembrane #status predicted <TMN>

F:303-315/Domain: intracellular #status predicted <INT>

F:37,69,112,161,254/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.7%; Score 82.5; DB 1; Length 315;

Best Local Similarity 23.8%; Pred. No. 4.2;

Matches 38; Conservative 22; Mismatches 43; Indels 57; Gaps 8;

QY 4 NVGONLVVDLSTQIFCHNDYPTITDYTLQRGSAVGVLSNFGSTVYKSGSSY-PF 52

Db 159 NINCSVFEIAT-----PEPITDNV-----EDHDTVTYTSDSINTVYAT 198

QY 53 SGSSYFPPTTSETPRVVYNSRDKPWPALYLTTPVSSAGLVKAGSLIAVLILRQTNV 112

Db 199 SGES-----TTDETPEPTIDKEEDHTVDTVGYSTVYTSSTGSL-----TTKS 240

QY 113 NSDDQFQVNNIYANNDDVVPPTG-GCDVSARDVYTLIPDR 151

Db 241 TTDDAD-LYDTYNDNDIVPSTTVGCS-----TTSISNWK 273

RESULT 19

H71456

probable pyrolysin (EC 3.4.-.-) homolog PH0310 [similarity] - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000

C:Accession: H71456

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Ogu

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A:Reference number: A71000; MUID:98344137; PMID:9679194

A:Accession: H71456

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1155 <KAW>

A:Cross-references: GB:AP000001; NID:93236128; PIDN:BAA29383.1; PID:93256700

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBa

C:Genetics:

A:Gene: PH0310

C:Keywords: hydrolase

Query Match 9.7%; Score 82.5; DB 2; Length 1155;

Best Local Similarity 24.5%; Pred. No. 21;

Matches 46; Conservative 24; Mismatches 59; Indels 59; Gaps 11;

QY 14 STQIFCHNDYPTITDYTLQRGSAVGVLSNFGSTVYKSGSSYPTTSETPR----- 67

Db 566 NTIWFHFN-YIDFWIDYIS--DKEYNAILSNLSLESYLPOAHFP--ESPAPDVQ 618

QY 68 ----VYV--NSRTDKPWPVALYLP-----VSSAGLVKAGSLIAVL 104

Db 619 LSDYLVYIMEDKNTPEVFTYVOPKDIIPDNFTGELYNTFLILSTIGGEYKVLGSL----- 674

QY 105 ILRQTNVNSDDFOFV-----VNNIYANNDDVVPPTGGCDVSARDVYTLIPDR--- 151

Db 675 ----EGNVSIPDGGYVVPLETPVVSVDV-APNNVTVQYGNFNYSISYI-IPLEDARIVY 728

QY 152 GSVPIPLT 159

Db 729 GNESTPPLT 736

RESULT 20

HMIVS3

hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)

C:Species: influenza A virus

C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998

C:Accession: B29971

R:Kida, H.; Shortridge, K.F.; Webster, R.G.

Virology 162, 160-166, 1988

A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in Chin

A:Reference number: A94370; MUID:88101364; PMID:3336940

A:Accession: B29971

Db 127 IVNTDSESTIDILSGSTHSPETSSKKPDYIDNSCNSVFEIATPEPTIDNV----- 178
QY 40 GGVLSNFGCTVKY-----SGSSYPPTTSETPRVYNSRTDKPWPVALYLPVS 88
Db 179 ----EDHTDVTYTSDSINTVSASSGES---TTDETPEPT-TDEEDHTVTDVTSYTTVS 229
QY 89 SAGGLVIKAGSLIAVLILROTNNYNSDDFQFVWNITIANNDVVPVT--GGCDVSARDVTVT 146
Db 230 TSSGIV-----TTKSTTDAD-LYDTYNDNDVPTTVGGS-----TTS 267
QY 147 LPDYR 151
Db 268 ISNRY 272

RESULT 24

hemagglutinin precursor - raccoonpox virus (strain WR)
C:Species: raccoonpox virus
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jul-1999
C:Accession: B43381
R:Cavallaro, K.F.; Esposito, J.J.
Virology 190, 434-439, 1992
A:Title: Sequences of the raccoon poxvirus hemagglutinin protein.
A:Reference number: A43381; MUID:92410621; PMID:1529542
A:Accession: B43381
A:Molecule type: DNA
A:Residues: 1-314 <CAV>
A:Cross-references: GB:M3956; NID:g335637; PIDN:AAA48252.1; PID:g335638
C:Superfamily: vaccinia virus hemagglutinin; Immunoglobulin homology
C:Keywords: glycoprotein; hemagglutinin; late protein; transmembrane protein
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-314/Product: hemagglutinin #status predicted <HEG>
F:27-105/Domain: immunoglobulin homology <IMM>
F:280-296/Domain: transmembrane #status predicted <TM>
F:34-103/Dsulfide bonds: #status predicted
F:37,69,112,161,253/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.4%; Score 80; DB 1; Length 314;
Best Local Similarity 22.7%; Pred. No. 7.1;
Matches 42; Conservative 21; Mismatches 48; Indels 74; Gaps 10;

QY 2 VVNWGQNLVVDLSIQIFCHN-----DY-----PETITDYVTLQSGSAY 39
Db 127 IVNTDSESTIDILSGSTHSPETSSKKPDYIDNSCNSVFEIATPEPTIDNV----- 178
QY 40 GGVLSNFGCTVKY-----SGSSYPPTTSETPRVYNSRTDKPWPVALYLPVS 88
Db 179 ----EDHTDVTYTSDSINTVSASSGES---TTDETPEPT-TDEEDHTVTDVTSYTTVS 229
QY 89 SAGGLVIKAGSLIAVLILROTNNYNSDDFQFVWNITIANNDVVPVT--GGCDVSARDVTVT 146
Db 230 TSSGIV-----TTKSTTDAD-LYDTYNDNDVPTTVGGS-----TTS 267
QY 147 LPDYR 151
Db 268 ISNRY 272

RESULT 25

neurokinin 3 receptor - mouse (fragment)
N:Alternate names: neuromedin K receptor; NK-3 receptor
C:Species: Mus musculus (house mouse)
C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 19-May-2000
C:Accession: S55524; I73045
R:Maroteaux, L.
submitted to the EMBL Data Library, June 1995
A:Reference number: S55524
A:Accession: S55524
A:Molecule type: mRNA
A:Residues: 1-385 <MAR>
A:Cross-references: EMBL:X87823; NID:g861055; PIDN:CAA61088.1; PID:g861056

R:Cook, G.A.; Elliott, D.; Metwali, A.; Blum, A.M.; Sandor, M.; Lynch, R.; Weinstock, J. Immunol. 152, 1830-1835, 1994
A:Title: Molecular evidence that granuloma T lymphocytes in murine schistosomiasis ma
A:Reference number: 156216; MUID:94165478; PMID:8120392
A:Accession: I73045
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 103-197, 'S', 199-266, 'P', 268-328 <COO>
A:Cross-references: GB:I27827; NID:g450288; PIDN:AAA17893.1; PID:g480780
C:Superfamily: neurokinin 1 receptor

Query Match 9.4%; Score 80; DB 2; Length 385;
Best Local Similarity 30.3%; Pred. No. 9.1;
Matches 30; Conservative 15; Mismatches 22; Indels 32; Gaps 6;

QY 27 ITDYVTLQSGSAYGGVLSNFGCTVKYSSYPPTTSETPRVYNSRTDK--PWPVAL 82
Db 30 ITEMIALQAG-----NFS-----SALGLPVTSGAPQVDRDNLTNQVQPSWRITAL 74
QY 83 YLTPVSSAGGLVIKA---GSLIAVLI-----LQOTNNY 112
Db 75 W----SLAVGLVAVAVFVGNLIVITILAHKRMKRTVNY 109

RESULT 26

hemagglutinin - vaccinia virus (strain Ankara)
C:Species: vaccinia virus
A:Variety: strain Ankara
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
C:Accession: T37438
R:Antoine, G.; Schefflinger, F.; Falner, F.G.; Dorner, F.
submitted to the EMBL Data Library, March 1997
A:Description: The complete genomic sequence of the Modified Vaccinia Ankara (MVA) st
A:Reference number: 220877
A:Accession: T37438
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-315 <ANT>
A:Cross-references: EMBL:U94848; PIDN:AAB96543.1
A:Experimental source: strain Ankara
C:Genetics:
A>Note: MVA165R
C:Superfamily: vaccinia virus hemagglutinin; Immunoglobulin homology

Query Match 9.4%; Score 79.5; DB 2; Length 315;
Best Local Similarity 22.2%; Pred. No. 7.9;
Matches 41; Conservative 21; Mismatches 50; Indels 73; Gaps 9;

QY 2 VVNWGQNLVVDLSIQIFCHN-----DY-----PETITDYVTLQSGSAY 39
Db 127 IVNTDSESTIDILSGSTHSPETSSKKPDYIDNSCNSVFEIATPEPTIDNV----- 178
QY 40 GGVLSNFGCTVKY-----SGSSYPPTTSETPRVYNSRTDKPWPVALYLPVS 88
Db 179 ----EDHTDVTYTSDSINTVSASSGES---TTDETPEPTTKEEDHTVTDVTSYTTVS 230
QY 89 SAGGLVIKAGSLIAVLILROTNNYNSDDFQFVWNITIANNDVVPVT--GGCDVSARDVTVT 146
Db 231 TSSGIV-----TTKSTTDAD-LYDTYNDNDVPTTVGGS-----TTS 268
QY 147 LPDYR 151
Db 269 ISNRY 273

RESULT 27

JH0155
pectin lyase (EC 4.2.2.10) D precursor - Aspergillus niger
C:Species: Aspergillus niger
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 21-Jul-2000
C:Accession: JH0155; P50295
R:Gysler, C.; Harmsen, J.A.M.; Kester, H.C.M.; Visser, J.; Heim, J.

Gene 89, 101-108, 1990
A:Title: Isolation and structure of the pectin lyase D-encoding gene from *Aspergillus niger*
A:Reference number: JH0155; MUID:90325592; PMID:2375363
A:Accession: JH0155
A:Molecule type: DNA
A:Residues: 1-373 <GYS1>
A:Cross-References: GB:M55657; GB:M30305; NID:g166515; PIDN:AAA32701.1; PID:g166516
A:Experimental source: Strain N756
A:Accession: PS0295
A:Molecule type: protein
A:Residues: 20-30 <GYS2>
C:Genetics:
A:Gene: pelD
A:Introns: 68/2; 116/2; 158/2; 190/3
C:Keywords: carbon-oxygen lyase; glycoprotein
F:1-19/Domains: signal sequence; status predicted <STC>
F:20-373/Product: pectin lyase D #status experimental <MAT>
F:128,274,348/Binding site: carboxylate (Asn) (covalent) #status predicted
F:128,274,348/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match	9.38;	Score 79;	DB 2;	Length 373;
Best Local Similarity	23.68;	Pred. No. 11;		
Matches 38;	Conservative 25;	Mismatches 72;	Indels 26;	Gaps 7

```
Oy      9 LVNDSTQIFCHANDYPETITDYVWLÖRGSAYGV-USNFSGIWKYSSGSEYPPPTSETPR 67  
         | :|::||| | | | | | :| :| :| :| :| :|  
Db     200 LGTDADSRVSTNNNYINGESDYSATCDGHHTWNVNVLDDGSSDKYTFCG-NLYLKFTSGRAPK 25
```

```
QY      68 V-----VYNSRTDKPWPVALYLTPVSAGGLVKAGSLIA-VLLRQTNN-----YNS 114
          |:::| : | : ::|| : : : : : : : : : : : : : : : : : : : : : : : :
Db     259 QODNTYLIHYNNYWENNSGHAPEI-----GSGGYLAEGNFIENVDTVLFTTTFEGALFSS 314
```

```

QY 115 DDFQFWNIYANDVYVPTGGCDVSARDVTV-----TLP 148
      | | | | |
Db 315 DSASTTSESYIGRSCAVANNGDGLGTSTTVLSNLSGDLP 355

```

RESULT 28

Probable protein kinase [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 Sequence Revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: B66465
 R:Thelloglis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
 Chli, C.W.; Chung, M.K.; Conn, L.F.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
 A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maltz, R.; Marziani
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salberg, S.L.; Schwartz, J.R.; Shum, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A6141; MUID:11016719; PMID:11150712
 A:Accession: B66465
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-1064 <SNO>
 A:Cross-references: GB:AE005172; NID:g10086466; PIDN:AG12526.1; GSPDB:GN0041
 C:Genetics:
 A:Map position: 1

Query Match	9.3%;	Score 79;	DB 2;	Length 1064;
Best Local Similarity	21.2%;	Pred. No. 40;		
Matches	45;	Conservative	21;	Mismatches 58;
				Indels 88;
				Gaps 10

```

0y      11 VDLSTQIACHNDYPETITDVTLQ-----RGSAYGV-----LSNFGTV 50
      :||| | | ||| : | | : |||| :
Db      602 IDLSYNNFTGN-IPETFSDLTQLQSLDSSNSLHGDIAYLGLSTSLASLINSNNNSGPI 666

```

```

QY      51 KYSGSSYFPFMTSETPRVYNSR-----TDKQWVALYLTPTSAG 91
          | | | | | : | :
Db      661 ---BSPFFKTIISTSYLONTNMLCHSLDITCSHTGONGVKSPKIVAL----- 707

```

```

Qy  92 GLVTKAGSLIAY-----LILKQTNYY-----NSDDFQEPWNIYANDDVVPTGK 135
      || | ||: ||| : | :||: | :|
Db 708 TAVILASITAILAAMDILIRNNHLYKTSQNSSSSPSTAEDESYWTF-----IPFQK 760

```

```
QY      136 CDVSARDVYVTLPD-----YRGSV 155
      ::::| |:::|
Db      761 LGITVNNIVTSLTDENVIGKSGIYKAEIP 792
```

RESULT 29

penicillin-binding protein dacF precursor dacF [imported] - Agrobacterium tumefaciens C:Species: Agrobacterium tumefaciens C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 C:Accession: AH3003 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo-erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClair, P.; Romero, P.; Zhang, S. Science 294, 2317-2323, 2001

ster, E.W.
A:Title: The Genome of the Natural Genetic Engineering Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193

A/Accession: AN0003
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-389 <KUR>

A:Cross-References: gb:EU008097; FIDN:ANL44440.1; FID:gt/742046; GSFD:UNU0107
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: dacF

A;Map position: linear chromosome

Query Match	9.3%	Score	78.5	DB	2	Length	389
Best Local Similarity	24.2%	Pred. No.	13				
Matches	43	Conservative	26	Mismatches	64	Indels	45
						Gaps	8

```
Oy      1 P V N V G N L W D I S T - Q I F C H N D ----- Y P E T I T D Y V T L Q R S A Y G V L S N F S G P V K Y S G   56
```



```
Db      47 P V A N A K M V D Y K T G K V I S H Q E A F R K W Y P A S L T K L M T ----- A Y I A F S O M K A G K I -----   97
```

```

QY 55 SSYPPTTSETPRVVYNSRTDKPWPVALY----LTPVSSAGGLVYKAGSLIAVLILRQT 109
    : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 98 -----SPQRELVMSKKAADQZASKMVFKPGQKITMDSALKLLIKSANDAVAI-AET 149

```

```

QY      110  NNYSNDEQFVNNIYAN-----NDVVPYTGCCDVSARDVT-----TLBDY 150
      11:|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      150  IGGTSDNFVAQMAAQARLGMSSTHYVNPNGIGKGGQYTTARDIALALALLIKREPEY 207

```

RESULT 30

penicillin-binding protein dacF precursor rp38 [imported] - *Agrobacterium tumefaciens*
C.Species: *Agrobacterium tumefaciens*
C.Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C.Accession: A99280
R.Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Gold
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A.Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium*
A.Reference number: A97359; PMID:11743194

```
A:Accession: A3200
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-398 <KUR>
```

A;Cisss references: GB:AE007070,
C;Genetics:
A;Gene: AGR_L_2387
A;Map position: linear chromosome

Query Match	9.38	Score	78.5	DB	2	Length	398
Best Local Similarity	24.28	Pred. No.	13				
Matches	43	Conservative	26	Mismatches	64	Indels	45
				Gaps	8		

Db 149 MGSSN-----ACKRGSGGFSFRLN--W-----LTKSGSTYPLVNTVMPNN 187
Oy 112 YNSDDQGFWMNYY---ANND-----VYVPTGGCDVSARDVTYTLPRDGSVP 155
Db 188 DNEFDKL-YTMGIHHPSTNOEQTSLYVOASGRVYSTRSQOTTIPNIGSRP 237

RESULT 41

D90176

hypothetical protein SS00334 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: D90176

R:She, O.; Stuph, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.

arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: D90176

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-433 <KUR>

A:Cross-References: GB:AE006641; NID:g13813479; PIDN:AAK40667.1; GSPDB:GN00155

C:Genetics:

A:Gene: SS00334

Query Match 9.1%; Score 77; DB 2; Length 433;
Best Local Similarity 22.7%; Pred. No. 20;
Matches 40; Conservative 22; Mismatches 42; Indels 72; Gaps 10;

Oy 24 PETITDYVTLQRCGSAVGVLNSFGVTKVSGSSYPP----- 60
Db 188 PNTWTSM-----TWFG---TFNGTA-YIGSKFYVDHMLYGVNVNQFQVYIMF 237
Oy 61 TSETPRVVYNSRTDKPMPVALY---LTPVS-----SAGGLYIKAGSLTAVILRQ 108
Db 238 ESSETP-----VAQTPKPKVQSVSVYGLPVPVNYSEINGTQGVAV-AGNMIV----- 286
Oy 109 TNNYNSDDQGFWMNYYANNDVYVPTGGCDVSARDVTYTLPRDGSVP--ITPLTV 161
Db 287 -----TIQPNKTIYTPSGNLSEVNFSEIVFYNPANNANPANGQVPLIAF 329

RESULT 42

T00467

Probable UDPglucose 4-epimerase (EC 5.1.3.2) [imported] - Arabidopsis thaliana

N:Alternate names: UDPgalactose 4-epimerase homolog

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Mar-2001

C:Accession: T00467; F84761

R:Roundley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul

submitted to the EMBL Data Library, April 1998

A:Description: Arabidopsis thaliana chromosome II BAC F1913 genomic sequence.

A:Reference number: Z14160

A:Accession: T00467

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-375 <ROU>

A:Cross-References: EMBL:AC004238; NID:g3033373; PID:g3033381

A:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Roundley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Mofatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: F84761

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-375 <STO>

A:Cross-References: GB:AE002093; NID:g3033381; PIDN:AAK12825.1; GSPDB:GN00139

C:Genetics:
A:Gene: F1913.8; At2g34850

A:Map position: 2

A:introns: 29/3; 103/3; 124/3; 203/2; 246/3; 295/1

C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homolo

C:Keywords: isomerase

Query Match 9.0%; Score 76.5; DB 2; Length 375;
Best Local Similarity 28.0%; Pred. No. 19;
Matches 28; Conservative 14; Mismatches 43; Indels 15; Gaps 3;

Oy 18 FCHNDYPTITDYVTLQRCGSAVGVLNSFGVTKVSGSSYPPPTSETPRVVYNSRTDKP 77
Db 125 YHNINISNTL---VLEETMAHVKVTLIYSSCATGCEPEKMPITETRYQV----- 172
Oy 78 WPVALYLPVSSAGGLVIT--KAGSLTAVILRQTNVNSD 115
Db 173 -PINPYGAKAKMAEDITLDFSKNSIMAVMILRFYFNIGSD 211

RESULT 43

J01154

hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragmen

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000

C:Accession: J01154

R:Yasuda, J.; Shortridge, K.F.; Shmiltzu, Y.; Kida, H.

J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian

A:Reference number: J01153; MUID:91341491; PMID:1875195

A:Accession: J01154

A:Molecule type: genomic RNA

A:Residues: 1-550 <YAS>

A:Cross-References: GB:D00930; NID:g221273; PIDN:BA00770.1; PID:g221274

A:Note: the authors translated the codon GGG for residue 218 as Gln

A:Note: residues 528-532 are not shown in this publication

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; hemotimer

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.0%; Score 76.5; DB 2; Length 550;
Best Local Similarity 24.3%; Pred. No. 30;
Matches 36; Conservative 18; Mismatches 59; Indels 35; Gaps 6;

Oy 21 NDYPTITDYVTLQRCGSAVGVLNS-----NFGTVKYSGSYPPTSETPRVVYNSRT 74
Db 96 NCPYVDPVYASLSRLVASSGTFLETFEGFTWGTGVQNGSN-----ACKRGPRANGFFSLR 151
Oy 75 DKRPWVALYLPVSSAGGLYIKAGSLTAVILRQTNVNSDQFQWNIY-----ANN 127
Db 152 N--W-----LTKSGSTYPLVNTVMPNNDNFDKL-YTMGVIHHPSTNOEQTN 193
Oy 128 DVVYPTGGCDVSARDVTYTLPRDGSVP 155
Db 194 LYVOASGRVYVSTRSQOTTIPNIGSRP 221

RESULT 44

J01155

hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000

C:Accession: J01155

R:Yasuda, J.; Shortridge, K.F.; Shmiltzu, Y.; Kida, H.

J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian

A:Reference number: J01153; MUID:91341491; PMID:1875195

A:Accession: J01155

A:Molecule type: genomic RNA

A;Residues: 1-550 <YAS>
 A;Cross-references: GB:D00931; NID:9221277; PIDN:BA00771.1; PID:9221278
 A;Note: the authors translated the codon GGC for residue 218 as Glu, GCC for residue 538
 A;Note: residues 528-532 are not shown in this publication
 C;Superfamily: Influenza virus haemagglutinin
 C;Keywords: glycoprotein; homotrimer
 F;1-32/Domain: haemagglutinin HA1 #status predicted <HA1>
 F;330-545/Product: haemagglutinin HA2 #status predicted <HA2>
 F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match

9.0%; Score 76.5; DB 2; Length 550;
 Best Local Similarity 24.3%; Pred. NO. 30;

Matches 36; Conservative 18; Mismatches 59; Indels 35; Gaps 6;

QY 21 NDYPETITDYTLRGSAVGGVLS-----NFGTVKSSGSSYPPTTSEPRVYNSRT 74
 Db 96 NCYDYDPDIASLSLVASSGTLEPTTEGFTWTCVTONGSN---ACKRGPAFGFSRL 151
 QY 75 DKPMPVALYLPVSSAGGLVIKAGSLIAVLLRQTNVNSDDPQFVWNIY-----ANN 127
 Db 152 N-W-----LTKSGSTYPLVLTWTPMNDNFCKL-YIMGVHHPTNOBOTN 193
 QY 128 DVVVPTGGCDVSARDVTVTLDPYRGVSP 155
 Db 194 LYVQASGRVTVSTRSQOTIIPNIGSRP 221

RESULT 45

HMLIVH

hemagglutinin precursor - Influenza A virus

C;Species: Influenza A virus

C;Date: 28-Feb-1981 #sequence.revision 28-Feb-1981 #text_change 22-Oct-1999

C;Accession: A93705; A93233; A04051; A93231; A94441

R;Both, G.W.; Sleight, M.J. 8, 2561-2575, 1980

Nucleic Acids Res. 8, 2561-2575, 1980

A;Title: Complete nucleotide sequence of the haemagglutinin gene from a human influenza

A;Reference number: A93705; MUID:81053698; PMID:6253883

A;Accession: A93705

A;Molecule type: genomic RNA

A;Residues: 1-566 <BOT>

A;Cross-references: GB:V01103

A;Experimental source: strain A/NT/60/68/29C

A;Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A/NT/68

R;Dopheide, T.A.; Ward, C.W.

FEBS Lett. 110, 181-183, 1980

A;Title: The disulphide bonds of a Hong Kong influenza virus haemagglutinin.

A;Reference number: A91276; MUID:80179105; PMID:6768586

A;Contents: annotation; disulfide bonds

R;Gething, M.J.; Bye, J.; Skehel, J.; Waterfield, M.

Nature 287, 301-306, 1980

A;Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from

A;Reference number: A93233; MUID:81030852; PMID:7421990

A;Accession: A93233

A;Molecule type: genomic RNA

A;Residues: 1-24, 'S', 26, 'D', 28-159, 'G', 161-197, 'I', 199-241, 'V', 243-249 <GET>

A;Experimental source: strain X-31[H3]

C;Superfamily: Influenza virus haemagglutinin

C;Keywords: haemagglutinin; homotrimer; lipoprotein; thiolester bond

F;1-16/Domain: signal sequence #status predicted <SIG>

F;17-344/Product: haemagglutinin HA1 #status predicted <HA1>

F;346-566/Product: haemagglutinin HA2 #status predicted <HA2>

F;536-552/Domain: transmembrane #status predicted <TM1>

F;30-482,68-293,80-92,155-489,297-321/Disulfide bonds: #status experimental

F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 9.0%; Score 76.5; DB 1; Length 566;
 Best Local Similarity 25.1%; Pred. NO. 31;

Matches 43; Conservative 22; Mismatches 67; Indels 39; Gaps 9;

QY 1 PYVAVGQNLVYDL---STQIRCHNDYPTETIDYTLRGSAVGGVLS-----NFGTVK 51
 Db 90 PHCDVFOETWDLFEVERKAF-SNCYPYDVPDIASLSLVASSGTLEPTTEGFTWTCVTQ 148

QY 52 YSGSSYPPTTSETPRVYNSRTDKPMPVALYLPVSSAGGLVIKAGSLIAVLLRQTN 111
 Db 149 NGGSN---ACKRGPDGFGFSRLN--W-----LTKSGSTYPLVLTWTPMNDNFCKL 167
 QY 112 YNSDDPQFVWNIY---ANNP---VVVPTGGCDVSARDVTVTLDPYRGVSP 155
 Db 188 DNFCKL-YIMGVHHPTNOBOTSIYVQASGRVTVSTRSQOTIIPNIGSRP 237

Search completed: November 28, 2002, 18:53:57
 Job time: 47 secs

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 499 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-399-886-3

Query Match 9.88; Score 83; DB 4; Length 499;
Best Local Similarity 28.58; Pred. No. 0.82;
Matches 39; Conservative 19; Mismatches 49; Indels 30; Gaps 8;

QY 7 QNLVVDLSTQIFCHNDP-----ETITDYVTLQSGSAVGSLSNFSQTV-----KXSGS 55
DB 229 QPLLYD-SIQFAAQRISFVLNANQTVGN-WVRANPFQTV--GFAGGINSALIRYQGA 284
QY 56 SYPEPTSETPRVYNSRTDKPMPVALYLPVSSAGGLVIKAGSLAVLILROTNNYNSD 115
DB 285 PVAEPTTQTPSVIPLIETNL-HPLARMPVPGSPTPGVDKA-----LNLAFNNGT 335
QY 116 DFOFWMNITVANDVVP 132
DB 336 NF-----FTNNATFTP 346

RESULT 5

US-09-396-260-3
Sequence 3, Application US/09396260
Patent No. 6184015
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 61840150 No. 6184015disk of No. 6184015th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/396,260
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/032,315
FILING DATE: 27-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 5200,200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 499 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-396-260-3

Query Match 9.88; Score 83; DB 4; Length 499;
Best Local Similarity 28.58; Pred. No. 0.82;
Matches 39; Conservative 19; Mismatches 49; Indels 30; Gaps 8;

QY 7 QNLVVDLSTQIFCHNDP-----ETITDYVTLQSGSAVGSLSNFSQTV-----KXSGS 55
DB 229 QPLLYD-SIQFAAQRISFVLNANQTVGN-WVRANPFQTV--GFAGGINSALIRYQGA 284
QY 56 SYPEPTSETPRVYNSRTDKPMPVALYLPVSSAGGLVIKAGSLAVLILROTNNYNSD 115
DB 285 PVAEPTTQTPSVIPLIETNL-HPLARMPVPGSPTPGVDKA-----LNLAFNNGT 335
QY 116 DFOFWMNITVANDVVP 132
DB 336 NF-----FTNNATFTP 346

RESULT 6

US-09-576-281-3
Sequence 3, Application US/09576281
Patent No. 6277611
GENERAL INFORMATION:
APPLICANT: Pedersen, Anders
APPLICANT: Svendsen, Allan
APPLICANT: Schneider, Palle
APPLICANT: Rasmussen, Grethe
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 62776110 No. 6277611disk of No. 6277611th America
STREET: 405 Lexington Avenue
CITY: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/576,281
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,318
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5032,200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 499 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-576-281-3

Query Match 9.88; Score 83; DB 4; Length 499;
Best Local Similarity 28.58; Pred. No. 0.82;
Matches 39; Conservative 19; Mismatches 49; Indels 30; Gaps 8;

QY 7 QNLVVDLSTQIFCHNDP-----ETITDYVTLQSGSAVGSLSNFSQTV-----KXSGS 55
DB 229 QPLLYD-SIQFAAQRISFVLNANQTVGN-WVRANPFQTV--GFAGGINSALIRYQGA 284
QY 56 SYPEPTSETPRVYNSRTDKPMPVALYLPVSSAGGLVIKAGSLAVLILROTNNYNSD 115
DB 285 PVAEPTTQTPSVIPLIETNL-HPLARMPVPGSPTPGVDKA-----LNLAFNNGT 335

OY 116 DQFVWNIYANDVVP 132
Db 336 NF-----FINNATFTP 346

RESULT 7

US-08-462-484-4
; Sequence 4, Application US/08462484
; Patent No. 5667531
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Xu, Feng
; APPLICANT: Dalboge, Henrik
; APPLICANT: Schneider, Palle
; APPLICANT: Aaslyng, Dorrit A.
; TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 56675310 No. 5667531disk of No. 5667531th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentln Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,484
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/441,147
; FILING DATE: 15-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowmey, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4185.010-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 878 9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 519 amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-462-484-4

Query Match 9.8%; Score 83; DB 1; Length 519;
Best Local Similarity 28.5%; Pred. No. 0.87;
Matches 39; Conservative 19; Mismatches 49; Indels 30; Gaps 8;

OY 7 ONLVLDSTOIFCHNDY-----ETIDYVTLRGSAVGVLSNFGSTV-----KYSGS 55
Db 249 QPLVLD-SIQIFAORSFVINAMOTVGNV-WVRANPNFGTV--GFAGINSAILIRYOGA 304
OY 56 SYPEPTSETPRVYNSRTDKPWPVALYLFVSSAGLVIRAGSLIAVLILROTNVNSD 115
Db 305 PVAERTTQTGTSVPIPLIFETNL-HPLARMYPVGSPTPGGVDA-----LNLAFNFGT 355
OY 116 DQFVWNIYANDVVP 132
Db 356 NF-----FINNATFTP 366

RESULT 8
US-08-441-147-4
; Sequence 4, Application US/08441147
; Patent No. 5770418

GENERAL INFORMATION:

APPLICANT: Yaver, Debbie Sue
APPLICANT: Xu, Feng
APPLICANT: Dalboge, Henrik
APPLICANT: Schneider, Palle
APPLICANT: Aaslyng, Dorrit A.
TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 57704180 No. 5770418disk of No. 5770418th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,147
FILING DATE: 15-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lowmey, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4185.010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 878 9655
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 519 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-441-147-4

Query Match 9.8%; Score 83; DB 1; Length 519;
Best Local Similarity 28.5%; Pred. No. 0.87;
Matches 39; Conservative 19; Mismatches 49; Indels 30; Gaps 8;

OY 7 ONLVLDSTOIFCHNDY-----ETIDYVTLRGSAVGVLSNFGSTV-----KYSGS 55
Db 249 QPLVLD-SIQIFAORSFVINAMOTVGNV-WVRANPNFGTV--GFAGINSAILIRYOGA 304
OY 56 SYPEPTSETPRVYNSRTDKPWPVALYLFVSSAGLVIRAGSLIAVLILROTNVNSD 115
Db 305 PVAERTTQTGTSVPIPLIFETNL-HPLARMYPVGSPTPGGVDA-----LNLAFNFGT 355
OY 116 DQFVWNIYANDVVP 132
Db 356 NF-----FINNATFTP 366

RESULT 9

PCT-US95-07536-4
; Sequence 4, Application PC/TUS9507536
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: PURIFIED POLYPORUS LACCASES AND
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

OY 142 DVTV 145
Db 94 DVTI 97

RESULT 12
US-08-169-927-2
; Sequence 2, Application US/08169927
; Patent No. 5783441
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell
; APPLICANT: Dobson, Michael E.
; APPLICANT: Ching, Wei Mei
; APPLICANT: Dasch, Gregory A
; TITLE OF INVENTION: Gene and Protein Applicable to the
; TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
; TITLE OF INVENTION: Rickettsia typhi and the Detection of both
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Counsel, Naval Medical R & D Command
; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,927
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/742,128
; FILING DATE: 08/09/91
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, A. David
; REGISTRATION NUMBER: 24,743
; REFERENCE/DOCKET NUMBER: 75,976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-1022
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1612 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-169-927-2

Query Match 8.8%; Score 75; DB 1; Length 1612;
Best Local Similarity 23.7%; Pred. No. 34;
Matches 31; Conservative 17; Mismatches 47; Indels 36; Gaps 5;

OY 31 VTLDGSAVYGV--LSNFSGTVKYSGSSYPPTSETPRVYNSRTDKPWFVALYLPVS 88
Db 908 ITVANLTYSIGRTKNNQGTLSG-----MPNNGRTY----- 943

OY 89 SAGLVIKAGLIVALLRQ---TNNINSDDPQFVWNIYANNVVVPTGGCDVSARDVTV 145
Db 944 ---GLIGENG---PKLKQVTFETDYNNLGSIIANNVTINDVTLTGGIAGIDFDPAKI 996

OY 146 TLPIYRGSVPI 156
Db 997 TLGSVNGNANV 1007

RESULT 13
US-07-937-609-21
; Sequence 21, Application US/07937609
; Patent No. 5319073

; GENERAL INFORMATION:
; APPLICANT: WANK, Stephen A.
; TITLE OF INVENTION: CLONING AND FUNCTIONAL EXPRESSION OF
; TITLE OF INVENTION: CHOLECYSTOKININ RECEPTOR-ENCODING DNA
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/937,609
; FILING DATE: 19920902
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/831,248
; FILING DATE: 07-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/861,769
; FILING DATE: 01-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/928,033
; FILING DATE: 11-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/166 NIHID
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; US-07-937-609-21

Query Match 8.6%; Score 73; DB 1; Length 411;
Best Local Similarity 28.3%; Pred. No. 7.8;
Matches 28; Conservative 16; Mismatches 23; Indels 32; Gaps 6;

OY 27 ITDVTLDGSAVYGVLSNFSGTVKYSGSSYPPTSETPRVYNSRTDK---PWPVAL 82
Db 30 VTEWLAIQAC-----NFS-----SALGLPATTQAFSQRANLTNCFVOPSWRIAL 74

OY 83 YLIPVSSAGLIVAKA---GSLIAVLI-----LRQTNVY 112
Db 75 W-----SLAYGLVAVAVAGNLIYIWIILAKRMKMTVINY 109

RESULT 14
US-08-029-170-21
; Sequence 21, Application US/08029170
; Patent No. 6169173
; GENERAL INFORMATION:
; APPLICANT: WANK, Stephen A.
; TITLE OF INVENTION: CLONING AND FUNCTIONAL EXPRESSION OF
; TITLE OF INVENTION: CHOLECYSTOKININ RECEPTOR-ENCODING DNA
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA

QY 132 PTGGCDVSARDVYTLPLD 149
DB 352 PTEILGSSLDNGTSLPLD 369

RESULT 20
US-09-295-186-11
Sequence 11, Application US/09295186B
Patent No. 6127137
GENERAL INFORMATION:
APPLICANT: Hasida, Miyoko
APPLICANT: Tsutsuimi, No. 61271371ko
APPLICANT: Halkier, Torben
APPLICANT: Strlinger, Mary Ann
TITLE OF INVENTION: An Acidic Phospholipase, Production, and
FILE REFERENCE: 4953.204-US
CURRENT APPLICATION NUMBER: US/09/295,186B
CURRENT FILING DATE: 1999-04-20
PRIOR APPLICATION NUMBER: 1215/96
PRIOR FILING DATE: 1996-10-31
PRIOR APPLICATION NUMBER: PCT/DK97/00490
PRIOR FILING DATE: 1997-10-30
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 11
LENGTH: 573
TYPE: PRT
ORGANISM: Hyphozyma sp. CBS 648.91
US-09-295-186-11

Query Match 8.3%; Score 70; DB 3; Length 573;
Best Local Similarity 23.9%; Pred. No. 27;
Matches 33; Conservative 16; Mismatches 45; Indels 44; Gaps 7;

QY 23 YPETITDYVLQ-----RCGAYGVV-----SNFGTVKXSGSSYPPTSEMPRVVYNS 72
DB 265 YQTSLDYFELSTADKTLNLSMGNKFSVEMSDVKNKSKTDSMPPI-----IIADE 318
QY 73 RTDKPMPVALYLPVSSAGGLVIRAGSLVLIROTNNYNSDDEQF--VNNIYANDVV 131
DB 319 R-----EPGLI---IPRNTIWEFNFEYERGSWN--PNVSAFI 351
QY 132 PTGGCDVSARDVYTLPLD 149
DB 352 PTEILGSSLDNGTSLPLD 369

RESULT 21
US-08-190-802A-53
Sequence 53, Application US/08190802A
Patent No. 5519003
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: P.O. Box 60850
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-0850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/190,802A

FILING DATE: 01-FEB-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 8600-0139
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: MSL1, Fig. 36
US-08-190-802A-53

Query Match 8.2%; Score 69.5; DB 1; Length 422;
Best Local Similarity 28.0%; Pred. No. 20;
Matches 26; Conservative 18; Mismatches 24; Indels 25; Gaps 6;

QY 59 FP---TTSETPRVYNSRT--DKPMPVALYLPVSSAGGLVIRAGSLVLIROTNNY 113
DB 52 FPLDITSDHRILSSFTSSQKPEDETIYISKIISLGH--IKWSL-----NNFD 100
QY 114 SDDQF-----VNNIYANDVV--PTGGCD 137
DB 101 MDEMFEKPNSTRFSPKHLVNDISIFFPNGECN 133

RESULT 22
US-08-190-802A-55
Sequence 55, Application US/08190802A
Patent No. 5519003
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: P.O. Box 60850
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-0850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/190,802A
FILING DATE: 01-FEB-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 8600-0139
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein

LENGTH: 422 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: MSL1, Fig. 36
US-08-487-072A-53

Query Match 8.2%; Score 69.5; DB 4; Length 422;
Best Local Similarity 28.0%; Pred. No. 20;
Matches 26; Conservative 18; Mismatches 24; Indels 25; Gaps 6;

QY 59 FP---TTSPTPRVYNSRT--DKPWPVALYLTPVSSAGGLVAKASLAVLLRQTNNVN 113
DB 52 FPDLLTTSDEHRTILSFTSSOKPDEITYISKISTLGH--IKWSSL-----NNFD 100
QY 114 SDDFOF-----VNNIYANDVVV--PTGGCD 137
DB 101 MDEMFKPENSTRFPKHLVNDISIFPNGECN 133

RESULT 28
US-08-487-072A-55
Sequence 55, Application US/08487072A
Patent No. 6423684
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Kon, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,072A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-0763
TELEFAX: (202) 887-1500

INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: ORF RB1, Fig. 38
US-08-487-072A-55

Query Match 8.2%; Score 69.5; DB 4; Length 422;
Best Local Similarity 28.0%; Pred. No. 20;
Matches 26; Conservative 18; Mismatches 24; Indels 25; Gaps 6;
QY 59 FP---TTSPTPRVYNSRT--DKPWPVALYLTPVSSAGGLVAKASLAVLLRQTNNVN 113

DB 52 FPDLLTTSDEHRTILSFTSSOKPDEITYISKISTLGH--IKWSSL-----NNFD 100
QY 114 SDDFOF-----VNNIYANDVVV--PTGGCD 137
DB 101 MDEMFKPENSTRFPKHLVNDISIFPNGECN 133

RESULT 29
US-09-331-581-23
Sequence 23, Application US/09331581
Patent No. 6130070
GENERAL INFORMATION:
APPLICANT: TONDA, Hideki
APPLICANT: HAMA, Yoko
APPLICANT: KUMAGAI, Hiromichi
TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
FILE REFERENCE: 0059-1142-0FCT
CURRENT APPLICATION NUMBER: US/09/331,581
CURRENT FILING DATE: 1999-06-30
EARLIER APPLICATION NUMBER: PCT/JP98/04929
EARLIER FILING DATE: 1998-10-30
EARLIER APPLICATION NUMBER: JP 9-314608
EARLIER FILING DATE: 1997-10-31
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 23
LENGTH: 332
TYPE: PRT
ORGANISM: Schwanniomycetes occidentalis
US-09-331-581-23

Query Match 8.1%; Score 69; DB 4; Length 332;
Best Local Similarity 28.8%; Pred. No. 16;
Matches 38; Conservative 12; Mismatches 48; Indels 34; Gaps 9;

QY 1 PVAVGNLVLDSTQIFCHNDYETITDYTLOR-----GSA--YGGYL-SNFGSTVK 51
DB 158 PVIDVSSNQFRD--PKVFWHERFKSMDCSEIARVXIOIFGSANLKNWLNSS--G 213
QY 52 YGSSYFPPTSETPRVYNSRTDKPWPVALYLTPVSSAGGLVAKASLAVLLRQTNN 111
DB 214 YYGQYQMSRLIEVP--IENSDSK-WVMLPLINPGSLPGST-----NQ 255
QY 112 Y--NSDDFOFV 120
DB 256 YFVGDFDGFQFV 267

RESULT 30
US-07-723-002C-4
Sequence 4, Application US/07723002C
Patent No. 5447862
GENERAL INFORMATION:
APPLICANT: Heim, Jutta
APPLICANT: Meyhack, Bernd
APPLICANT: Gysler, Christof
APPLICANT: Visser, Jacob
APPLICANT: Kester, Hermanus Cornelis Maria
TITLE OF INVENTION: No. 5447862el Expression System
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/723,002C
; FILING DATE: 28-JUN-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8702475
; FILING DATE: 04-FEB-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 150,880
; FILING DATE: 29-JAN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8818046.8
; FILING DATE: 28-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8914666.6
; FILING DATE: 26-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 384,898
; FILING DATE: 24-JUL-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Foley, Shawn P.
; REGISTRATION NUMBER: 33,071
; REFERENCE/DOCKET NUMBER: 4-16317/+CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8615
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 379 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-723-002C-4

Query Match
; Best Local Similarity 8.1%; Score 69; DB 1; Length 379;
; Matches 34; Conservative 21; Mismatches 69; Indels 26; Gaps 6;

QY 21 NQPTTIDVTLQKRGATGCVLSNFGSGSTPFTTSTPRVYNSRDKPMV 80
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 213 NNVIDGSDYSATCGHYGIYLDGADLVTKNGNIYHTSGSPVQDNTLH---CV 269
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 81 ALVLTPTS-----SAGGLYIKAGSLI---AVLLRQTNMNSDDQF-----VMNI 123
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 270 NNFYDISGAFETGEGGIVLAEQNVQNDVYL-----ETLEGATVFTSTAGEVCS 324
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 124 YANNVVVPTGGCD-VSARDVTVLTPDYRG 152
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 325 YLGRDCVINGFGSGSFSEDSSTFLSDPFG 354
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 31
; US-08-712-072C-2
; Sequence 2, Application US/08712072C
; Patent No. 5925541
; GENERAL INFORMATION:
; APPLICANT: Jack Goldstein, Alex Zhu and Lin Leng
; TITLE OF INVENTION: ENDO-BETA-GALACTOSIDASE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rochstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/712,072C
; FILING DATE: 11-SEP-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bogosian, Elizabeth A.
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 63475/97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELE: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; US-08-712-072C-2
```

```

Query Match
; Best Local Similarity 8.1%; Score 69; DB 2; Length 422;
; Matches 48; Conservative 23; Mismatches 67; Indels 74; Gaps 11;
```

```

QY 12 DLSTQIFCHND-----YPTETIDVTLQKRGATGCVLSN- 45
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 214 NLVTMTSPNDIRFYVNSLQITYARVSGGCTQOMPDPVPTLLNAGGAGMPGAIITNA 273
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 46 ---FSGTVKY-----SGSSYFPFTTSTPRVYNS-----RTDKPM 78
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 274 DLFPMSQVDVRYVKLPFGNSGDESGVIYPMWTWGGSSVSDARTGKIRETGCT 333
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 79 PVALYLTPTS-----SAGGL-VIKAGSLAVLLRQTNMNSD--DFQVWNIYANNV 129
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 334 SIEQYLTGLTPNTTYRPGGAKVASAQSISIV---KNYGTAVDATIGTTSYSNNSV 389
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 130 VVPTGGCDVSARDVTVLTPDYRGVPIPLAVY 161
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 390 TPTTGG-----ANNITATVYFK---PLSGIVY 413
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

RESULT 32
; US-09-620-412C-309
; Sequence 309, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 309
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
; US-09-620-412C-309
```

```

Query Match
; Best Local Similarity 8.1%; Score 69; DB 4; Length 619;
; Matches 27; Conservative 21; Mismatches 61; Indels 18; Gaps 5;
QY 31 VTLQKRSATGCVLSNFGSGSTPFTTSTPRVYNSR---TDKPMVAL 82
```

Db 402 ITFEGNSAVHGAIYKNGVLFEFLGNAGPLAEKENTTIANGCAITYSNFKANQGSPLIF 461
QY 83 YLTPVSSAGGLVIRKAGSLIAYLILQTNNNYSDDFQFVWNIYANDVYPTGCGVNSARD 142
Db 462 SONHANKKGG-----AIIYAQVYNEQMODTIRFKNTAKEGGAIITSSQCSITAHN 512
QY 143 VYVTLPD 149
Db 513 -TITFSD 518

RESULT 33
US-09-071-035-198
; Sequence 198, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071.035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8512
; TELEFAX: (301) 309-8504
; INFORMATION FOR SEQ ID NO: 198:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-198

Query Match 8.1%; Score 68.5; DB 4; Length 129;
Best Local Similarity 24.7%; Pred. No. 4.6;
Matches 20; Conservative 10; Mismatches 30; Indels 21; Gaps 3;

QY 18 FCHNDYPTITDYYTLORGSAVGVLSNFGTVKXSGSSYPPTTSEPRVYNSRTDKP 77
Db 16 FCHSD-----TAFGEAAVNSGVSPFYG-TYEPTEESTTATSNSTTTP 60
QY 78 WPAVLYLTPVSSAGGLVIKAG 98
Db 61 -----TKPADGASVYLSG 75

RESULT 34
US-09-216-295-20
; Sequence 20, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:

; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328el Variant Egitr-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216.295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Emericella desertoru
US-09-216-295-20

Query Match 8.1%; Score 68.5; DB 4; Length 246;
Best Local Similarity 21.9%; Pred. No. 12;
Matches 35; Conservative 20; Mismatches 60; Indels 45; Gaps 6;

QY 38 AVGVLSNFGTVKXSGSYP-----PPTSEPRVYNSRT-----DKWPVALYLT 85
Db 88 SYANAQFTSTIKLSLSSIPSMKMOYSTDIYANVAIDLFTSSAGGDSIEIETMILA 147
QY 86 PYSSAGGLVIRKAGSLIAYLIL-----ROTNNYSDDFQFVWN 122
Db 148 ALGAGP-ISSFGSLATVTLGVTWLSLQPNQSMQYVSFVASTTESFSADLMDFI-N 205
QY 123 IYANDVY-----VPTGCDVSARDVYVTLPDYRGSV 154
Db 206 YLAENQGLSSQSYLTHVQAGTEPFGTDATLTVSSVSV 245

RESULT 35
US-09-032-523-2
; Sequence 2, Application US/09032523
; Patent No. 6232454
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032.523
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0479 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: RATRNOT02
CLONE: 947429
US-09-032-523-2

Query Match 8.1%; Score 68.5; DB 4; Length 415;
Best Local Similarity 19.7%; Pred. No. 25;
Matches 35; Conservative 20; Mismatches 56; Indels 67; Gaps 6;

QY 12 DLSQJFCHNDYETITDVTYLRGSAYGVLNFGSGTVKSGSSYPPTSETPRVYN 71
DB 201 DVERDNCRDIYAVN-----GGEVNDARIGICGDSPPAPIVSERNELIQ 249
QY 72 -----SRTDKP---WPVALYLP-VSSAGGLVIRKAGSLIA 102
DB 250 FLSDLSLTADGFIGHYIFRPKLPPTTEGPVTTTFPTTGKLPVALCQCKRRTGTL-- 307
QY 103 VLLRGTNNVNSDFQF-----VMNIYANDVYVPFGGCDYSAR 141
DB 308 -----EGNYCSDPVLACTVITTTTRDGLHATVSIINIKGKGLAIQAGKMSAR 359

RESULT 36
US-08-947-965-73
Sequence 73, Application US/08947965A

Patent No. 6004790
GENERAL INFORMATION:
APPLICANT: Dijkhuizen, Lubbert
APPLICANT: Andersen, Carsten
APPLICANT: Ostlen, Claus von der
TITLE OF INVENTION: Cyclomaltodextrin glucanotransferase
FILE REFERENCE: 4285.204-US
CURRENT APPLICATION NUMBER: US/08/947,965A
CURRENT FILING DATE: 1997-10-09
EARLIER APPLICATION NUMBER: 0477/95
EARLIER FILING DATE: 1995-04-21
EARLIER APPLICATION NUMBER: 1173/95
EARLIER FILING DATE: 1995-10-17
EARLIER APPLICATION NUMBER: 1281/95
EARLIER FILING DATE: 1995-11-16
EARLIER APPLICATION NUMBER: PCT/DK96/00179
EARLIER FILING DATE: 1996-04-22
NUMBER OF SEQ ID NOS: 78
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 73
LENGTH: 686
TYPE: PRT
ORGANISM: Bacillus sp.
US-08-947-965-73

Query Match 8.1%; Score 68.5; DB 3; Length 686;
Best Local Similarity 24.5%; Pred. No. 51;
Matches 26; Conservative 20; Mismatches 37; Indels 23; Gaps 4;

QY 2 VVANGONLVVDLSIQICNDYETITDVTYLRGSAY---GGVLNFGSGTVKSGSSY 57
DB 431 VVALNRN-----NTPASITGLVSLRASVNDVGLGLNGNTLVGAGGAAS 478
QY 58 PPTTSETPRVYNSRDKPPAPVALYLPVSSAGGLVIRKAGSLAV 103
DB 479 NF-TLAPGTAVMOYTTDATPI-----IGNVSPMAKPGVITTI 517

RESULT 37
5194600-2
Patent No. 5194600
APPLICANT: BUSSEY, HOWARD; BOONE, CHARLES; SOMMER, STEVE S.;
HILL, KATHRYN; MEADEN, PHILIP

TITLE OF INVENTION: GENES WHICH PARTICIPATE IN B-GLUCAN
ASSEMBLY AND USE THEREOF
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,316
FILING DATE: 05-MAR-1990
SEQ ID NO: 2
LENGTH: 291
5194600-2

Query Match 8.0%; Score 68; DB 6; Length 291;
Best Local Similarity 26.2%; Pred. No. 17;
Matches 39; Conservative 23; Mismatches 67; Indels 20; Gaps 7;

QY 12 DLSQJFCHNDYETITDVTYLRGSAYGVLNFGSGTVKSGSSYPPTSETPRVYN 71
DB 109 DLSQJFCHNDYETITDVTYLRGSAYGVLNFGSGTVKSGSSYPPTSETPRVYN 160
QY 72 SRTDKPVALYLPVSSAGGLVIRKAGSLIAVILRQ--TNNY---NSDDFVNIYAN 126
DB 161 TGT-RPDPSTDETPVSA---VTSLSIDSYITTEGTTSTVTTTRAPTSMMVTVVROG 215
QY 127 NDVVPFGGCD-VSARDVYLPDYGVS 154
DB 216 NITVQTFVQRRSSQYVTVASPSV-GSI 243

RESULT 38
US-09-071-035-244
Sequence 244, Application US/09071035

Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis polynucleotides and polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 244:
SEQUENCE CHARACTERISTICS:
LENGTH: 624 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-071-035-244

Query Match 8.0%; Score 68; DB 4; Length 624;
Best Local Similarity 21.1%; Pred. No. 51;
Matches 42; Conservative 24; Mismatches 61; Indels 72; Gaps 8;


```

: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/071.035
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: A. Anders Brookes
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: P3369P2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 242:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1301 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-071-035-242

Query Match      8.0%; Score 68; DB 4; Length 1301;
Best Local Similarity 21.1%; Pred. No. 1.5e+02;
Matches 42; Conservative 24; Mismatches 61; Indels 72; Gaps 8;

QY 11 VDLSTQIRCHNDYPT-----ITDVTUQ-----RGS----- 37
   ||||| :||| :|
Db 959 VDLATIGSFDDYDEYTPTKILRLVKDSKGDITTFITSDPAKGTVYXSAKDPQAF 1018
   ||||| :||| :|

QY 38 --AYGG-----VLNFSGTVKYSGSSYPPTTSETPRVYNSRTDKPMP---VAL 82
   ||||| :||| :|
Db 1019 ILAXGGQELRVTLPTKVKADVSGDYNSAEQNTFGQRKIKTIVY--NIIPVXKKKDVYI 1076
   ||||| :||| :|

QY 83 YLTPVSSAGGLVYKAG-----SLIAVLLIRQTNNNNSDDPQVWNIIYA 125
   ||||| :||| :|
Db 1077 KVGKQSGNGATIKLGEXFEFTSSDIPAEYAGVEMSSIDRLVDXHKDFSGQWSVFA 1136
   ||||| :||| :|

QY 126 NNDVVPFG-----GCDYS 139
   ||||| :||| :|
Db 1137 NSNFVLADGTVKVNGDDIS 1155
   ||||| :||| :|

RESULT 44
US-08-612-734B-2
: Sequence 2, Application US/08612734B
: Patent No. 5914246
: GENERAL INFORMATION:
: APPLICANT: Peery, Robert B.
: APPLICANT: Skatrud, Paul L.
: APPLICANT: Tobin, Matthew B.
: TITLE OF INVENTION: Multiple Drug Resistance Gene of
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Eli Lilly and Company
: STREET: Lilly Corporate Center, DC1501
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: U.S.A.
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/612.734B
: FILING DATE: 08-MAR-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Craig, Anne I.
: REGISTRATION NUMBER: 32,976
: REFERENCE/DOCKET NUMBER: X-9681
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-354-9570
: TELEFAX: 617-354-4043
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1349 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-612-734B-2

Query Match      8.0%; Score 68; DB 2; Length 1349;
Best Local Similarity 25.3%; Pred. No. 1.6e+02;
Matches 42; Conservative 17; Mismatches 61; Indels 46; Gaps 7;

QY 3 VNVGQNLVVDLST-----QIFCHNDYPTITPYTLORGSAY-----GGV 42
   ||||| :||| :|
Db 371 VNVGQVLTVLMSILIGSRSLGNVAPNGCAFT--NGVAAAKIYTIIDRRSPLDPIYSDECKV 429
   ||||| :||| :|

QY 43 LSNFSGTVKYSGSSYPPTTSE-----TPRVYNSRTDKPMPVAL---YL 84
   ||||| :||| :|
Db 430 LDHFEQNIIEFRNVKHIVPSRPEVTVMEDVSLSPMAGKTTLALVGPSGSGKSTYVGLVERFY 489
   ||||| :||| :|

QY 85 TPVSAGGLVTKAGSLIAVLILR---QTNNYNSDDDPQVWNIIYAN 126
   ||||| :||| :|
Db 490 LPV---GGQVLDDHDIDTLNLRWLROQISLVSOEPLVFTTIPRN 532
   ||||| :||| :|

RESULT 45
US-09-320-878-2
: Sequence 2, Application US/09320878A
: Patent No. 6117659
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanie C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
: APPLICANT: TANG, Li
: TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
: FILE REFERENCE: 300622002120
: CURRENT APPLICATION NUMBER: US/09/320.878A
: CURRENT FILING DATE: 1999-05-27
: EARLIER APPLICATION NUMBER: CIP OF 09/141,908
: EARLIER FILING DATE: 1998-08-28
: EARLIER APPLICATION NUMBER: CIP OF 09/073,538
: EARLIER FILING DATE: 1998-05-06
: EARLIER APPLICATION NUMBER: CIP OF 08/846,247
: EARLIER FILING DATE: 1997-04-30
: EARLIER APPLICATION NUMBER: 60/119,139
: EARLIER FILING DATE: 1999-02-08
: EARLIER APPLICATION NUMBER: 60/100,880
: EARLIER FILING DATE: 1998-09-22
: EARLIER APPLICATION NUMBER: 60/087,080
: EARLIER FILING DATE: 1998-05-28
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 2
: TYPE: PRT
: LENGTH: 3739
: ORGANISM: Streptomyces venezuelae
: US-09-320-878-2

Query Match      8.0%; Score 68; DB 3; Length 3739;
Best Local Similarity 23.4%; Pred. No. 6.8e+02;
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 17:21:36 ; Search time 19 seconds
(without alignments)
134.937 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186

Perfect score: 848
Sequence: 1 PVTNNGNMLVVDLSTQIFCH.....DYTVTLPDYRGSVPIPLTVY 161

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_Aa.*
1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep.*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	279	10	US-09-900-575-29
2	845	99.6	279	10	US-09-900-575-36
3	838	98.8	279	10	US-09-900-575-42
4	837	98.7	279	10	US-09-900-575-55
5	834	98.3	279	10	US-09-900-575-28
6	834	98.3	279	10	US-09-900-575-32
7	834	98.3	279	10	US-09-900-575-34
8	834	98.3	279	10	US-09-900-575-44
9	834	98.3	300	10	US-09-912-020-367
10	833	98.2	279	10	US-09-900-575-37
11	832	98.1	279	10	US-09-900-575-27
12	831	98.0	279	10	US-09-900-575-23
13	831	98.0	279	10	US-09-900-575-43
14	829	97.8	279	10	US-09-900-575-38
15	829	97.8	279	10	US-09-900-575-40
16	827	97.5	279	10	US-09-900-575-39
17	827	97.5	280	10	US-09-900-575-30
18	826	97.4	279	10	US-09-900-575-25
19	826	97.4	279	10	US-09-900-575-26

20	826	97.4	279	10	US-09-900-575-45	Sequence 45, Appl
21	825	97.3	279	10	US-09-900-575-31	Sequence 31, Appl
22	823	97.1	279	10	US-09-900-575-35	Sequence 35, Appl
23	819	96.6	279	10	US-09-900-575-24	Sequence 24, Appl
24	819	96.6	279	10	US-09-900-575-33	Sequence 33, Appl
25	819	96.6	279	10	US-09-900-575-41	Sequence 41, Appl
26	83	9.8	499	10	US-09-732-350-3	Sequence 3, Appl1
27	75.5	8.9	358	10	US-09-815-242-12939	Sequence 12939, A
28	75.5	8.9	448	10	US-09-815-242-5806	Sequence 5806, Ap
29	75.5	8.9	451	10	US-09-815-242-13127	Sequence 13127, A
30	71	8.4	410	10	US-09-804-551B-18	Sequence 18, Appl
31	70.5	8.3	471	10	US-09-881-752A-18	Sequence 18, Appl
32	70.5	8.3	511	10	US-09-815-242-13697	Sequence 13697, A
33	69	8.1	184	9	US-09-741-233A-6	Sequence 6, Appl1
34	69	8.1	185	10	US-09-970-638-5	Sequence 5, Appl1
35	69	8.1	245	10	US-09-995-598-24	Sequence 24, Appl
36	69	8.1	619	10	US-09-841-132-309	Sequence 309, Appl
37	69	8.1	5701	10	US-09-864-761-37319	Sequence 37319, A
38	69	8.1	5795	10	US-09-815-242-12610	Sequence 12610, A
39	68.5	8.1	239	10	US-09-995-598-50	Sequence 50, Appl
40	68.5	8.1	244	10	US-09-995-598-12	Sequence 12, Appl
41	68.5	8.1	415	9	US-09-905-291A-104	Sequence 104, Appl
42	68.5	8.1	415	10	US-09-909-320-104	Sequence 104, Appl
43	68.5	8.1	415	10	US-09-909-088B-104	Sequence 104, Appl
44	68.5	8.1	922	10	US-09-886-468-19	Sequence 19, Appl
45	68	8.0	3739	9	US-09-860-846-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-09-900-575-29
Sequence 29, Application US/09900575
Patent No. US20020150587A1
GENERAL INFORMATION:
APPLICANT: Langermann, Solomon
APPLICANT: Revel, Andrew
APPLICANT: Auguste, Christine
APPLICANT: Burlein, Jeanne
TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
FILE REFERENCE: 469201-549
CURRENT APPLICATION NUMBER: US/09/900,575
CURRENT FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: US/60/216,750
PRIOR FILING DATE: 2000-07-07
NUMBER OF SEQ ID NOS: 64
SOFTWARE: PatentIn version 3.0
SEQ ID NO 29
LENGTH: 279
TYPE: PRT
ORGANISM: E. coli
US-09-900-575-29

Query Match 100.0%; Score 848; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.8e-83;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PVTNNGNMLVVDLSTQIFCHNDYPTETDVTYTLQRGSAVYSGVSNFSGYKYSSTSPFP 60
Db PVTNNGNMLVVDLSTQIFCHNDYPTETDVTYTLQRGSAVYSGVSNFSGYKYSSTSPFP 85
QY 61 TTSETPVTNNSRTRDKWPALYLTPTVSSAGGLYKAGSLIAVILRQTNNNYSDDPQFV 120
Db TTSETPVTNNSRTRDKWPALYLTPTVSSAGGLYKAGSLIAVILRQTNNNYSDDPQFV 145
QY 121 WNIYANDVVPYTGCDVSARDVTVTLPDYRGSVPIPLTVY 161
Db WNIYANDVVPYTGCDVSARDVTVTLPDYRGSVPIPLTVY 186
RESULT 2
US-09-900-575-36


```

Qy      3  VVVGONLVVDLSTQIFCHNDYPETITDYVTLQRGSAAGGVLSNFGSYVSGSSYPFPT 62
Db      28  VVVGONLVVDLSTQIFCHNDYPETITDYVTLQRGSAAGGVLSNFGSYVSGSSYPFPT 87

Qy      63  SETPRVYVNSRTDKPMPVALYLPVSSAGGLYIKAGSLIAYLILKQTNVNSDDFOFVWN 122
Db      88  SETPRVYVNSRTDKPMPVALYLPVSSAGGVYIKAGSLIAYLILKQTNVNSDDFOFVWN 147

Qy      123  IYANDVYVPTGGCDVSARDVYTLPLDYGSPILPLTY 161
Db      148  IYANDVYVPTGGCDVSARDVYTLPLDYGSPILPLTY 186

RESULT 6
US-09-900-575-32
; Sequence 32, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Buriel, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 32
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-32

Query Match      98.3%; Score 834; DB 10; Length 279;
Best Local Similarity 98.1%; Pred. No. 5,5e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0

Qy      1  PAVNVGONLVVDLSTQIFCHNDYPETITDYVTLQRGSAAGGVLSNFGSYVSGSSYPFP 60
Db      26  PAVNVGONLVVDLSTQIFCHNDYPETITDYVTLQRGSAAGGVLSNFGSYVSGSSYPFP 85

Qy      61  TTSETPRVYVNSRTDKPMPVALYLPVSSAGGLYIKAGSLIAYLILKQTNVNSDDFOFV 120
Db      86  TTSETPRVYVNSRTDKPMPVALYLPVSSAGGVYIKAGSLIAYLILKQTNVNSDDFOFV 145

Qy      121  IYANDVYVPTGGCDVSARDVYTLPLDYGSPILPLTY 161
Db      146  IYANDVYVPTGGCDVSARDVYTLPLDYGSPILPLTY 186

RESULT 7
US-09-900-575-34
; Sequence 34, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Buriel, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 279

```

[illegible]

```

1  APPLICANT: XU, H. Howard
2  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
3  TITLE OF INVENTION: ESCHERICHIA COLI
4  FILE REFERENCE: ELITRA..001DVI
5  CURRENT APPLICATION NUMBER: US/09/912,020
6  CURRENT FILING DATE: 2001-07-23
7  PRIOR APPLICATION NUMBER: 09/492,709
8  PRIOR FILING DATE: 2000-01-27
9  PRIOR APPLICATION NUMBER: 60/117,405
10 PRIOR FILING DATE: 1999-01-27
11 NUMBER OF SEQ ID NOS: 485
12 SOFTWARE: FastSeq for Windows Version 3.0
13 SEQ ID NO 367
14 LENGTH: 300
15 TYPE: PR1
16 ORGANISM: E. Coll
17 US-09-912-020-367

```

Query Match	98.3%	Score 834;	DB 10;	Length 300;
Best Local Similarity	98.1%	Pred. No. 6.1e-82;		
Matches 158;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0;

QY	1	PVAVAGGNLAVDLSTOJLECHNDXPEETIDVYTLORGASVGGVLSNEGCTAKYGGSSVFP	60
Db	47	PVAVAGGNLAVDLSTOJLECHNDXPEETIDVYTLORGASVGGVLSNEGCTAKYGGSSVFP	1068
QY	61	TTSTSPRYVYNSRFDKWPALALYTPVSSAGGLYIKAGSLAVALLKQTNNYNSDFOFV	120
Db	107	TTSTSPRYVYNSRFDKWPALALYTPVSSAGGVAIKAGSLAVALLKQTNNYNSDFOFV	166
QY	121	WNIIYANDVYVPTGGCDVSARDVYVTLPLDYGVSPIRLAYV	161
Db	167	WNIIYANDVYVPTGGCDVSARDVYVTLPLDYGVSPIRLAYV	207

```

RESULT 10.
US-09-900-575-37
; Sequence 37, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burteln, Jeanne
; TITLE OF INVENTION: Fifth Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 37
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
; US-09-900-575-37

```

Query Match	98.2%	Score 833;	DB 10;	Length 279;
Best Local Similarity	97.5%	Pred. No. 7.1e-82;		
Matches 157; Conservative	2;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	1	PVVVGGGNLVVLDSTQIFEGCHNDPETHDVTVLQFGSAAGCVGLSNISGVIKAYGSSGSPF	60
	26	PAVVGGNVLVVDLSTQIFEGCHNDPETHDVTVLQFGSAAGCVGLSNISGVIKAYGSSGSPF	85
Db			
Qy	61	TTSTPTPVVYVNSRDKRWPAALYITPVSSAGGLYIKAGSLIAVLILRQTNNVMSDFQV	120
	86	TTSTPTPVVYVNSRDKRWPAALYITPVSSAGGLYIKAGSLIAVLILRQTNNVMSDFQV	145
Db			
Qy	121	WNIYANDVYVPTGGCDVSARDYVTVLPDVRGSPVLPILFVY	161
	146	WNIYANDVYVPTGGCDVSARDYVTVLPDVRGSPVLPILFVY	186
Db			
Qy			

RESULT 11
 US-09-900-575-27
 Sequence 27, Application US/09900575
 Patent No. US20020150587A1
 GENERAL INFORMATION:
 APPLICANT: Langermann, Solomon
 APPLICANT: Revel, Andrew
 APPLICANT: Auguste, Christine
 APPLICANT: Burlein, Jeanne
 TITLE OF INVENTION: Fifth Adhesin Proteins and Methods of Use
 FILE REFERENCE: 469201-549
 CURRENT APPLICATION NUMBER: US/09/900,575
 CURRENT FILING DATE: 2001-07-06
 PRIOR APPLICATION NUMBER: US/60/216,750
 PRIOR FILING DATE: 2000-07-07
 NUMBER OF SEQ ID NOS: 64
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 27
 LENGTH: 279
 TYPE: PRT
 ORGANISM: E. coli
 US-09-900-575-27

Query Match	98.1%;	Score 832;	DB 10;	Length 279;
Best Local Similarity	97.5%;	Pred. No. 9.1e-82;		
Matches 157;	Conservative 1;	Mismatches 3;	Indels 0;	Gaps 0;

Qy	1	PVYVANGOLVYDLS	STQJLEFCHNDHDEE	ETITDVTYTL	LORGSAAGCVLS	NFSGCYTK	SSSSS	PP	60
Db	26	PAVYANGOLVYDLS	STQJLEFCHNDHDEE	ETITDVTYTL	LORGSAAGCVLS	NFSGCYTK	SSSSS	PP	85
Qy	61	TTSEPRVRYVNSR	DKPAPVLYLT	PVSSAGGLY	IKRAGSL	IAVLTIL	QCTNNYNSD	DFQV	120
Db	86	TTSEPRVRYVNSR	DKPAPVLYLT	PVSSAGGLY	IKRAGSL	IAVLTIL	QCTNNYNSD	DFQV	145
Qy	121	WNIYANDVVPV	PTGCGDVSAD	VTYTL	LPDYGVS	VPIDLYY			161
Db	146	WNIYANDVVPV	PTGCGDVSAD	VTYTL	LPDYGVS	VPIDLYY			186

```

RESULT 12
US-09-900-575-23
Sequence 23, Application US/09900575
Patent No. US20020150587A1
GENERAL INFORMATION:
APPLICANT: Langemann, Solomon
APPLICANT: Revel, Andrew
APPLICANT: Auguste, Christine
APPLICANT: Barlein, Jeanne
TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
FILE REFERENCE: 469201-549
CURRENT APPLICATION NUMBER: US/09/900,575
CURRENT FILING DATE: 2001-07-06
PRIORITY APPLICATION NUMBER: US/60/216,750
PRIORITY FILING DATE: 2000-07-07
NUMBER OF SEQ ID NOS: 64
SOFTWARE: PatentIn version 3.0
SEQ ID NO 23
LENGTH: 279
TYPE: PRT
ORGANISM: E. coli
US-09-900-575-23

```

	Query Match	98.0%	Score 831	DB 10:	Length 279;
	Best Local Similarity	Pred.	No. 1.2e-81;		
Matches	157; Conservative	1;	Mismatches 3;	Indels 0;	Gaps 0;
OY	1 PVVANGQLVVLDSTOLFCHNDYPETITDVTYLQSGAYGGLSNFSGTAVKSYGSSEYFP	60			
b0	26 PVVANGQLVVLDSTOLFCHNDYPETITDVTYLQSGAYGGLSNFSGIVRSYGSSGYFP	85			

SOFTWARE: PatentIn version 3.0

```

; SEQ ID NO 39
; LENGTH: 279
; TYPE: prt
; ORGANISM: E. coli
US-09-900-575-39

```

Query Match	97.5%;	Score 827;	DB 10;	Length 279;
Best Local Similarity	96.9%;	Pred. No. 3.1e-81;		
Matches 156; Conservative	1;	Mismatches 4;	Indels 0;	Gaps 0;

QY	1	PVVNAGOLVYDLSIOJFCHNDYETITDYVTLORGSAAGGVLSNFGYKVSSSYP	60
Db	26	PAVNVGNLVYDLSIOJFCHNDYETITDYVTLORGSAAGGVLSNFGYKVSSSYP	85
QY	61	TTSETPRVYVNSRDKPMPVALYLTVPSSAGGLVTKKGLAVLLKQTNVNSDDDFQV	120
Db	86	TTSETPRVYVNSRDKPMPVALYLTVPSSAGGVAIKKGLAVLLKQTNVNSDDDFQV	145
QY	121	WNIYANDVYVPTGGCDVSAADYVTLTPDYGVSGPILFLTYV	161
Db	146	WNIYANDVYVPTGGCDVSAADYVTLTPDYGVSGPILFLTYV	186

```

RESULT 17
US-09-900-575-30
Sequence 30, Application US/09900575
Patent No. US20020150587A1
GENERAL INFORMATION:
APPLICANT: Langemann, Solomon
APPLICANT: Revel, Andrew
APPLICANT: Auguste, Christine
APPLICANT: Burteln, Jeanne
TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
FILE REFERENCE: 469201-549
CURRENT APPLICATION NUMBER: US/09/900,575
CURRENT FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: US/60/216,750
PRIOR FILING DATE: 2000-07-07
NUMBER OF SEQ ID NOS: 64
SOFTWARE: PatentIn version 3.0
SEQ ID NO 30
LENGTH: 280
TYPE: PRT
ORGANISM: E. coli
US-09-900-575-30

```

[illegible]

RESULT 18
 US-09-900-575-25
 ; Sequence 25, Application US/09900575
 ; Patent No. US20020150587A1
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: Langermann, Solomon
 ; APPLICANT: Revel, Andrew
 ; APPLICANT: Auguste, Christine
 ; APPLICANT: Birllein, Jeanne
 ;
 ; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use

```

: FILE REFERENCE: 469201-549
:
: CURRENT APPLICATION NUMBER: US/09/900,575
:
: CURRENT FILING DATE: 2001-07-06
:
: PRIOR APPLICATION NUMBER: US/60/216,750
:
: PRIOR FILING DATE: 2000-07-07
:
: NUMBER OF SEQ ID NOS: 64
:
: SOFTWARE: PatentIn version 3.0
:
: SEQ ID NO 25
:
: LENGTH: 279
:
: TYPE: prt
:
: ORGANISM: E. coli
:
: US-09-900-575-25

```

[illegible]

```

RESULT 19
US-09-900-575-26
? Sequence 26, Application US/09900575
? Patent No. US20020150587A1
? GENERAL INFORMATION:
? APPLICANT: Langemann, Solomon
? APPLICANT: Revel, Andrew
? APPLICANT: Auguste, Christine
? APPLICANT: Burlein, Jeanne
? TITLE OF INVENTION: Ftmh Adhesin Proteins and Methods of Use
? FILE REFERENCE: 469201-549
? CURRENT APPLICATION NUMBER: US/09/900,575
? CURRENT FILING DATE: 2001-07-06
? PRIOR APPLICATION NUMBER: US/60/216,750
? PRIOR FILING DATE: 2000-07-07
? NUMBER OF SEQ ID NOS: 64
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 26
? LENGTH: 279
? TYPE: PRT
? ORGANISM: E. coli
? US-09-900-575-26

```

```

Query Match 97.4%; Score 826; DB 10; Length 279;
Best Local Similarity 96.3%; Pred. No. 4e-81;
Matches 155; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAVNVGQNLVYDLSLTQLPCHNDYPETITIDYVTLQGSAYGGVLSNFSGTVKYSGSSYFPP 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26 PAVNVGQNLVYDLSLTQLPCHNDYPETITIDYVTLQGSAYGGVLSNFSGTVKYSGSSYFPP 85

QY 61 TTSETPRVYVNSRDKPMPVALLTLPVSSAGGLYKAGSLIAVLLRQTNNTNSDDFOV 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 86 TTSETPRVYVNSRDKPMPVALLTLPVSSAGGLYKAGSLIAVLLRQTNNTNSDDFOV 145

QY 121 WNIYANDVYVPTGGCDYSARDVYVTLDPDYGSVPIPLTV 161
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 146 WNIYANDVYVPTGGCDYSARDVYVTLDPDYGSVPIPLTV 186

```

RESULT 20
US-09-900-575-45
; Sequence 45, Application US/09900575

QY	113	NSDDF	117
	:	111	
db	202	SVDDF	206

```

RESULT 33
US-09-741-233A-6
: Sequence 6, Application US/09741233A
: Patent No. US20020157142A1
: GENERAL INFORMATION:
: APPLICANT: Glover, Lyn W.
: APPLICANT: Budziszewski, Gregory J.
: APPLICANT: Zhou, Qing
: APPLICANT: Levin, Joshua Z.
: TITLE OF INVENTION: Herbicide Target Genes and Methods
: FILE REFERENCE: PB/5-31127A
: CURRENT APPLICATION NUMBER: US/09/741,233A
: CURRENT FILING DATE: 2000-12-18
: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 6
: LENGTH: 184
: TYPE: PRT
: ORGANISM: Arabidopsis thaliana
US-09-741-233A-6

```

```

RESULT 34
US-09-970-638-5
: Sequence 5, Application US/09970638
: Patent No. US2002009062A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc.
: APPLICANT: Meyers, Rachel
: TITLE OF INVENTION: 27419, A NOVEL HUMAN ARGININE-N-METHYL
: TITLE OF INVENTION: TRANSFERASE AND USES THEREOF
: FILE REFERENCE: 38155-20039.00
: CURRENT APPLICATION NUMBER: US/09/970,638
: CURRENT FILING DATE: 2002-03-12
: PRIOR APPLICATION NUMBER: US 60/237,717
: PRIOR FILING DATE: 2000-10-05
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5
: LENGTH: 185
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Consensus amino acid sequence
US-09-970-638-5

```

```

Dd      93 GMLAFTAMD---VYFTSKRCHPIPEAID-KPFGFSGSPPTHHKQITFYMEDPLV 148
Qy      96 KAGSLI--AVLILQTNNNYNSDDFQFWNIYANNDVVVPT 133
Dd      149 KTGDTIEGSTITSKNNKKNNRDLDTLSMNVNGROD---PT 185

```

```

RESULT 35
US-09-995-598-24
: Sequence 24, Application US/09995598
: Patent No. US20020142293A1
: GENERAL INFORMATION:
: APPLICANT: CRAINIC, RADU
: APPLICANT: CARO, VALERIE
: APPLICANT: GUILLOT, SOPHIE
: TITLE OF INVENTION: ENTEROVIRAL POLYNUCLEOTIDES, METHODS OF DETECTING ENTEROVIRUSE
: TITLE OF INVENTION: CONTAINING THE POLYNUCLEOTIDES
: FILE REFERENCE: 215059US-660-660-0
: CURRENT APPLICATION NUMBER: US/09/995, 598
: CURRENT FILING DATE: 2001-11-29
: PRIOR APPLICATION NUMBER: US 60/253741
: PRIOR FILING DATE: 2000-11-29
: NUMBER OF SEQ ID NOS: 96
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 24
: LENGTH: 245
: TYPE: PRN
: ORGANISM: Echovirus
US-09-995-598-24

```

```

RESULT 36
US-09-841-132-309
: Sequence 309, Application US/09841132
: Patient No. US20070061848A1
: GENERAL INFORMATION:
: APPLICANT: Bhalla, Ajay
: APPLICANT: Skelky, Yasir A.W.
: APPLICANT: Probst, Peter
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
: TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
: FILE REFERENCE: 210121.469C8
: CURRENT APPLICATION NUMBER: US/09/841,132
: CURRENT FILING DATE: 2001-04-23
: NUMBER OF SEQ ID NOS: 599
: SOFTWARE: FastSeq for Windows Version 3.0/4.0
: SEQ ID NO 309
: LENGTH: 619
: TYPE: PRT
: ORGANISM: Chlamydia trachomatis
US-09-841-132-309

```



```

1      RESULT 41
2      US-09-905-291A-104
3      Sequence 104, Application US/09905291A
4      Patent No. US20020160374A1
5      GENERAL INFORMATION:
6      APPLICANT: Genentech, Inc.
7      APPLICANT: Ashkenazi, Avi
8      APPLICANT: Botstein, David
9      APPLICANT: Desnoyers, Luc
10     APPLICANT: Eaton, Dan L.
11     APPLICANT: Ferrara, Napoleone
12     APPLICANT: Filvaroff, Ellen
13     APPLICANT: Fong, Sherman
14     APPLICANT: Gao, Wei-Qiang
15     APPLICANT: Gerber, Hanspeter
16     APPLICANT: Gertlisen, Mary E.
17     APPLICANT: Goddard, A.
18     APPLICANT: Godowski, Paul J.
19     APPLICANT: Grimaldi, Christopher J.
20     APPLICANT: Gurney, Austin L.
21     APPLICANT: Hillan, Kenneth J.
22     APPLICANT: Kljavin, Ivar J.
23     APPLICANT: Mather, Jennie P.
24     APPLICANT: Pan, James
25     APPLICANT: Paoni, Nicholas F.
26     APPLICANT: Roy, Margaret Ann
27     APPLICANT: Stewart, Timothy A.
28     APPLICANT: Tunas, Daniel
29     APPLICANT: Williams, P. Mickey
30     APPLICANT: Wood, William, I.
31     TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
32     FILE REFERENCE: 10466-14
33     CURRENT APPLICATION NUMBER: US/09/905,291A
34     CURRENT FILING DATE: 2001-07-12
35     PRIOR APPLICATION NUMBER: PCT/US00/04414
36     PRIOR FILING DATE: 2000-02-22
37     PRIOR APPLICATION NUMBER: US 60/143,048
38     PRIOR FILING DATE: 1999-07-07
39     PRIOR APPLICATION NUMBER: US 60/145,698
40     PRIOR FILING DATE: 1999-07-26
41     PRIOR APPLICATION NUMBER: US 60/146,222
42     PRIOR FILING DATE: 1999-07-28
43     PRIOR APPLICATION NUMBER: PCT/US99/20594
44     PRIOR FILING DATE: 1999-09-08
45     PRIOR APPLICATION NUMBER: PCT/US99/20944
46     PRIOR FILING DATE: 1999-09-13
47     PRIOR APPLICATION NUMBER: PCT/US99/21090

```



```
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Geo, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertelsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paonli, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Thomas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/909,088B
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 104
LENGTH: 415
TYPE: PRT
ORGANISM: Homo sapiens
US-09-909-088B-104

Query Match      8.1%: Score 68.5; DB 10; Length 415;
Best Local Similarity 19.7%: Pred. No. 10;
Matches 35; Conservative 20; Mismatches 56; Indels 67; Gaps 6;
```

12 DLSOLFCHNDYPTITTYVTLGRSAYGVLNFSGTGKYSYSGSSYPPPTTSEPRVYVN 71

```
Db 201 DVERDNYCRYDYAVFN-----||::: ||||| |||||
Qy 72 -----SRTDKP-----WPAALVLP-VSSAGLVIRAGSLIA 102
Db 250 FLDSLSTADGFLGHYFRPKKLPPTTEQPVYTTFTGLKPVVALCOQKCRRTGL-- 307
Qy 103 VLLRQTNNSYSDPFCF-----VNNIYANDVVPYTGCGDVSAR 141
Db 308 -----EGNVCSSDFVLAGTVITTTIRDSGLHATVSTINITYKGNLAIQAGKMSAR 359

RESULT 44
US-09-886-468-19
Sequence 19, Application US/09886468
Patent No. US20020037293A1
GENERAL INFORMATION:
APPLICANT: Aventis Pasteur Limited
TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses th
FILE REFERENCE: 77813-5
CURRENT APPLICATION NUMBER: US/09/886,468
CURRENT FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 60/113,280
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,281
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,282
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,283
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,284
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,285
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,385
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/114,050
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,056
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,057
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,058
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,059
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,061
PRIOR FILING DATE: 1998-12-28
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 922
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-886-468-19

Query Match      8.1%: Score 68.5; DB 10; Length 922;
Best Local Similarity 27.4%: Pred. No. 30;
Matches 31; Conservative 15; Mismatches 34; Indels 33; Gaps 6;
```

53 SGSSYPPP--TTSEPRVYNSRTDKPMPVALYTPVS-SAGGLVIRAGSLIANLIRQT 109

Db 264 SGSSTPVPIVTFSDNKGIFERNHSINGGALIVARKLISSGGFTL----- 309

Qy 110 NNYSDDGFYVWNI-YANDV-----VVVPTGG-CDVSARDVYVTLPLDPRGSSVP 155

Db 310 -----FINNISYANSQNLGGAIAIDTGTETLSAEKGTITFGQNRSLP 353

RESULT 45
US-09-860-846-33
Sequence 33, Application US/09860846
Patent No. US20020164742A1

```
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: xue, Y.
; APPLICANT: zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-33
```

```
Query Match      8.0%; Score 68; DB 9; Length 3739;
Best Local Similarity 23.4%; Pred. No. 2.2e+02;
```

```
Matches 41; Conservative 25; Mismatches 61; Indels 48; Gaps 8;
```

```
OY 24 PETITDYVTLQGSAYGCVLSNFSGTVKYSGSSYPF---PTSETPRV-VYNSRTDKPW 78
|||:| ||:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
DB 2348 PETVGLGTLRRDNGQHRITLSLAEMAWANGLTYDMSILPTTTHPDLPYAFQTERYW 2407
OY 79 PVALILTPVSSAGLVYKA-----GSLIIVLILROTNNYNSD--- 115
|||:| ||:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
DB 2408 P-----QPLSLAAGDITSGAGAEHPILGAVALADSDGCLITGSLRTHPWLADHAV 2462
OY 116 -----DFOFVWNITYANDVVVPTGGCDVSAR--DVTVTLPDYRGSVPIPLTV 160
||:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
DB 2463 ACTVLLPGTAFVELAFRAGDQV---GCDLVEELTLDAPLVLP-RRGAVRVOLSV 2512
```

```
Search completed: November 28, 2002, 19:01:24
Job time : 23 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 18:54:01 ; Search time 66 Seconds
(without alignments)
325.051 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 161
Sequence: 1 PNVNQGNIWVLDLSTQIFCH.....DVTVTLPDPRGSRVPIPLTVY 161

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

A.Geneseq_101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	161	100.0	279	23	AAE18419
2	92	57.1	268	23	ABG68154
3	92	57.1	279	22	ABAB47074
4	92	57.1	279	22	AAAB72839
5	92	57.1	279	22	AAV72515
6	92	57.1	279	23	AAE18424
7	92	57.1	279	23	AAE18426
8	92	57.1	279	23	AAE18433
9	92	57.1	296	22	AAAB47073
10	92	57.1	300	16	AAAB76745

11	92	57.1	300	21	AAAB16009	E. coli proliferat
12	92	57.1	300	21	AAV59456	E. coli PC31 FlmH
13	92	57.1	300	23	AAU077488	Immunogenic comple
14	92	57.1	304	22	AAAB47072	Escherichia coli s
15	90	55.9	279	23	AAE18417	Escherichia coli s
16	90	55.9	279	23	AAE18418	Escherichia coli s
17	90	55.9	279	23	AAE18422	Escherichia coli s
18	90	55.9	279	23	AAE18429	Escherichia coli s
19	90	55.9	279	23	AAE18432	Escherichia coli s
20	90	55.9	279	23	AAE18436	Escherichia coli F
21	90	55.9	300	16	AAAB76769	FlmH protein deriv
22	90	55.9	300	16	AAAB76771	FlmH protein deriv
23	90	55.9	300	16	AAAB76773	FlmH protein deriv
24	90	55.9	300	16	AAAB76774	FlmH protein deriv
25	90	55.9	300	16	AAAB76763	FlmH protein deriv
26	89	55.3	296	16	AAAB76767	FlmH protein deriv
27	79	49.1	300	16	AAAB76772	FlmH protein deriv
28	75	46.6	280	23	AAE18420	Escherichia coli s
29	67	41.6	279	23	AAE18415	Escherichia coli s
30	67	41.6	279	23	AAE18416	Escherichia coli s
31	67	41.6	279	23	AAE18425	Escherichia coli s
32	67	41.6	279	23	AAE18427	Escherichia coli s
33	67	41.6	279	23	AAE18430	Escherichia coli s
34	67	41.6	279	23	AAE18431	Escherichia coli s
35	67	41.6	279	23	AAE18434	Escherichia coli s
36	67	41.6	279	23	AAE18435	Escherichia coli s
37	60	37.3	279	23	AAE18428	Escherichia coli s
38	59	36.6	300	16	AAAB76768	FlmH protein deriv
39	57	35.4	300	16	AAAB76766	FlmH protein deriv
40	56	34.8	279	23	AAE18423	Escherichia coli s
41	56	34.8	300	16	AAAB76765	FlmH protein deriv
42	56	34.8	300	16	AAAB76770	FlmH protein deriv
43	56	34.8	408	23	ABAB09458	Fusion protein deriv
44	48	29.8	279	23	AAE18413	Escherichia coli s
45	46	28.6	300	16	AAAB76764	FlmH protein deriv
46	43	26.7	279	23	AAE18414	Escherichia coli s
47	43	26.7	279	23	AAE18421	Escherichia coli s
48	43	26.7	300	16	AAAB76775	FlmH protein deriv
49	43	26.7	300	16	AAAB76776	FlmH protein deriv
50	38	23.6	131	21	AAV44760	E. coli type I pil
51	38	23.6	227	21	AAV44761	E. coli type I pil
52	35	21.7	35	22	AAV72511	Novel protein comp
53	32	19.9	94	22	AAV72513	Novel protein comp
54	31	19.3	31	22	AAV72510	Novel protein comp
55	31	19.3	110	22	AAV72514	Novel protein comp
56	22	13.7	22	22	AAV72509	Novel protein comp
57	22	13.7	71	22	AAV72512	Novel protein comp
58	21	13.0	21	22	AAV72518	Novel protein comp
59	21	13.0	21	22	AAV72520	Novel protein comp
60	19	11.8	19	22	AAV72519	Novel protein comp
61	16	9.9	18	23	ABG68161	Optimised FlmH adh
62	16	9.9	39	23	ABG68238	Optimised FlmH adh
63	15	9.3	21	23	ABG68241	Optimised FlmH adh
64	13	8.1	20	22	AAV72517	MBD-2-TQIF peptide
65	9	5.6	9	22	AAV72508	Mannose binding do
66	9	5.6	11	23	ABG68160	Optimised FlmH adh
67	9	5.6	12	23	ABG68240	Optimised FlmH adh
68	9	5.6	32	23	ABG68235	Optimised FlmH adh
69	9	5.6	32	23	ABG68236	Optimised FlmH adh
70	9	5.6	270	23	AAU091291	Human NOV6 protein
71	9	5.6	275	22	AAE08592	Human NOV13 protei
72	9	5.6	285	22	AAE08590	Human NOV11 protei
73	9	5.6	290	22	AAE08587	Human NOV8 protei
74	8	5.0	39	23	ABG68237	Optimised FlmH adh
75	8	5.0	719	22	AAAB8513	Human protein kina
76	8	5.0	1021	22	ABAB61000	Human protein kina
77	8	5.0	1036	23	ABAB80923	Novel human protei
78	7	4.3	9	23	ABG68159	Optimised FlmH adh
79	7	4.3	30	23	ABG68233	Optimised FlmH adh
80	7	4.3	30	23	ABG68234	Optimised FlmH adh
81	7	4.3	36	21	AAAG16565	Arabidopsis thalia
82	7	4.3	96	21	AAAG48964	Arabidopsis thalia
83	7	4.3	100	21	AAAG16564	Arabidopsis thalia

84	7	4.3	100	21	Aradidopsis thalia	157	6	3.7	40	22	AAW57484	Human brain expres
85	7	4.3	114	21	Human ORFX ORF268	158	6	3.7	40	22	AAW69879	Human bone marrow
86	7	4.3	114	23	Human ORF2096, prot	159	6	3.7	40	22	AAW17705	Peptide #4139 enco
87	7	4.3	117	21	Human secreted pro	160	6	3.7	40	22	AAW30219	Peptide #4256 enco
88	7	4.3	119	21	Aradidopsis thalia	161	6	3.7	40	22	AAW05367	Peptide #4049 enco
89	7	4.3	119	21	Aradidopsis thalia	162	6	3.7	40	23	ABG39509	Human peptide enco
90	7	4.3	120	21	Aradidopsis thalia	163	6	3.7	43	22	ABG13080	Novel human diagno
91	7	4.3	120	21	Aradidopsis thalia	164	6	3.7	43	22	ABG20394	Novel human diagno
92	7	4.3	120	21	Aradidopsis thalia	165	6	3.7	46	22	ABG65012	Human secreted pro
93	7	4.3	120	21	Aradidopsis thalia	166	6	3.7	47	21	AAW34623	Human secreted pro
94	7	4.3	122	21	Human polypeptide	167	6	3.7	50	22	AAW35585	Antiangiogenic pen
95	7	4.3	132	21	Aradidopsis thalia	168	6	3.7	52	22	AAW62515	Proionibacterium
96	7	4.3	135	21	Aradidopsis thalia	169	6	3.7	54	22	ABW28377	Peptide #1028 enco
97	7	4.3	136	19	Human myelin p2 pr	170	6	3.7	54	22	ABW33555	Peptide #1061 enco
98	7	4.3	136	21	Aradidopsis thalia	171	6	3.7	54	22	ABW19012	Protein #1011 enco
99	7	4.3	149	21	Aradidopsis thalia	172	6	3.7	54	22	AAW54329	Human brain expres
100	7	4.3	149	21	Aradidopsis thalia	173	6	3.7	54	22	AAW66729	Human bone marrow
101	7	4.3	160	21	Zea mays protein f	174	6	3.7	54	22	AAW62460	Human immune/haema
102	7	4.3	185	23	Streptococcus poly	175	6	3.7	54	22	AAW14595	Peptide #1029 enco
103	7	4.3	185	23	Streptococcus poly	176	6	3.7	54	22	AAW27016	Peptide #1053 enco
104	7	4.3	193	23	Streptococcus poly	177	6	3.7	54	22	AAW02320	Peptide #1002 enco
105	7	4.3	197	21	Eucalyptus grandis	178	6	3.7	54	23	ABG36383	Human peptide enco
106	7	4.3	210	23	Lactococcus lactis	179	6	3.7	57	22	AAW68010	Proionibacterium
107	7	4.3	245	21	Aradidopsis thalia	180	6	3.7	60	21	AAW25494	Aradidopsis thalia
108	7	4.3	245	21	Aradidopsis thalia	181	6	3.7	60	21	AAW15151	Plasmin inhibitor
109	7	4.3	260	21	Aradidopsis thalia	182	6	3.7	60	22	AAW22804	Human prostate can
110	7	4.3	260	23	Aradidopsis thalia	183	6	3.7	60	22	AAW96110	Human reproductive
111	7	4.3	261	21	Hericicidally activ	184	6	3.7	61	21	AAW03117	Human secreted pro
112	7	4.3	263	22	Aradidopsis thalia	185	6	3.7	61	22	AAW51020	Early 3/6, 7K prote
113	7	4.3	274	21	C glutamicum prote	186	6	3.7	62	22	ABW29727	Peptide #2378 enco
114	7	4.3	276	21	Aradidopsis thalia	187	6	3.7	62	22	ABW34899	Peptide #2405 enco
115	7	4.3	283	21	Aradidopsis thalia	188	6	3.7	62	22	ABW20316	Protein #2315 enco
116	7	4.3	283	21	Aradidopsis thalia	189	6	3.7	62	22	AAU22679	Novel human colon
117	7	4.3	286	21	Aradidopsis thalia	190	6	3.7	62	22	AAW55709	Human brain expres
118	7	4.3	304	21	Aradidopsis thalia	191	6	3.7	62	22	AAW68087	Human bone marrow
119	7	4.3	304	21	Aradidopsis thalia	192	6	3.7	62	22	AAW93125	Human digestive sy
120	7	4.3	304	22	Aradidopsis thalia	193	6	3.7	62	22	AAW15911	Peptide #2345 enco
121	7	4.3	349	21	Putative P. abyssi	194	6	3.7	62	22	AAW28416	Peptide #2453 enco
122	7	4.3	397	22	Bacteriophage 192	195	6	3.7	62	22	AAW03647	Peptide #2329 enco
123	7	4.3	397	22	Drosophila melanog	196	6	3.7	62	23	ABG37614	Human peptide enco
124	7	4.3	443	23	Lactococcus lactis	197	6	3.7	69	23	ABP03120	Human ORFX protein
125	7	4.3	628	15	Human H13 viral re	198	6	3.7	71	21	AAW25493	Aradidopsis thalia
126	7	4.3	628	15	Human H13 viral re	199	6	3.7	71	21	AAW45137	Proionibacterium
127	7	4.3	629	13	Full-length human	200	6	3.7	73	22	ABW303076	Novel human diagno
128	7	4.3	629	15	Human H13 viral re	201	6	3.7	75	23	AAU75822	Rice l. endoxylanase
129	7	4.3	629	15	Human H13 viral re	202	6	3.7	78	23	AAU21275	Human novel foetal
130	7	4.3	629	15	Human H13 viral re	203	6	3.7	78	23	ABP05923	Human ORFX protein
131	7	4.3	629	15	Human H13 viral re	204	6	3.7	81	22	ABW10553	Human pancreatic c
132	7	4.3	629	15	Human H13 viral re	205	6	3.7	81	22	AAW22089	Human digestive sy
133	7	4.3	629	15	Human H13 viral re	206	6	3.7	82	20	AAW07954	Human secreted pro
134	7	4.3	629	15	Human H13 viral re	207	6	3.7	82	21	AAW42148	Human secreted pro
135	7	4.3	629	15	Human H13 viral re	208	6	3.7	83	21	AAW15145	Protein of Txlin 1
136	7	4.3	629	20	Human retroviral r	209	6	3.7	83	21	AAW15146	Protein of Txlin 2
137	7	4.3	722	20	N. meningitidis st	210	6	3.7	83	21	AAW15147	Protein of Txlin 3
138	7	4.3	722	20	Proionibacterium	211	6	3.7	83	21	AAW15148	Protein of Txlin 4
139	7	4.3	997	21	Aradidopsis thalia	212	6	3.7	83	21	AAW15149	Protein of Txlin 5
140	7	4.3	1000	21	Aradidopsis thalia	213	6	3.7	83	21	AAW15150	Protein of Txlin 6
141	7	4.3	1001	21	Aradidopsis thalia	214	6	3.7	83	22	AAW53165	Proionibacterium
142	7	4.3	1004	21	Aradidopsis thalia	215	6	3.7	83	23	ABP05774	Human ORFX protein
143	7	4.3	1011	21	Aradidopsis thalia	216	6	3.7	91	21	AAW25492	Aradidopsis thalia
144	7	4.3	1015	21	Aradidopsis thalia	217	6	3.7	92	21	ABW39169	Staphylococcus epi
145	7	4.3	1040	21	Aradidopsis thalia	218	6	3.7	93	21	AAW32710	Eucalyptus grandis
146	7	4.3	1040	22	Aradidopsis thalia	219	6	3.7	93	21	ABW29270	Peptide #1921 enco
147	7	4.3	1040	22	Aradidopsis thalia	220	6	3.7	99	22	ABW34440	Peptide #1946 enco
148	7	4.3	1061	22	Aradidopsis thalia	221	6	3.7	99	22	ABW19850	Protein #1849 enco
149	7	4.3	3263	22	Drosophila melanog	222	6	3.7	99	22	AAW55226	Human brain expres
150	7	3.7	24	21	Human secreted pro	223	6	3.7	99	22	AAW67622	Human bone marrow
151	7	3.7	31	21	Human secreted pro	224	6	3.7	99	22	AAW15428	Peptide #1862 enco
152	7	3.7	31	21	Human secreted pro	225	6	3.7	99	22	AAW27914	Peptide #1951 enco
153	7	3.7	31	22	Human colon cancer	226	6	3.7	99	23	AAW03188	Peptide #1870 enco
154	7	3.7	40	22	Peptide #4168 enco	227	6	3.7	99	23	ABG37203	Human peptide enco
155	7	3.7	40	22	Peptide #4235 enco	228	6	3.7	101	22	ABW39316	Corynebacterium gl
156	7	3.7	40	22	Protein #4059 enco	229	6	3.7	103	23	AAW17722	Human lipocalin-2

376	6	3.7	270	22	AAB67838	Amino acid sequenc	449	6	3.7	358	21	AAG25356	Arabidopsis thalia
377	6	3.7	271	22	AAU87132	Novel central nerv	450	6	3.7	360	21	AAG09035	Arabidopsis thalia
378	6	3.7	271	23	AAG66039	Scfv 1b4 antibody	451	6	3.7	360	21	AAG19661	Arabidopsis thalia
379	6	3.7	272	21	AAB24140	Cucumber antimicro	452	6	3.7	361	21	AAG09034	Arabidopsis thalia
380	6	3.7	272	22	AAG91844	C glutamincum prote	453	6	3.7	361	21	AAG19660	Arabidopsis thalia
381	6	3.7	273	21	AAG39879	Arabidopsis thalia	454	6	3.7	362	22	AAE01937	Arabidopsis thalia
382	6	3.7	273	22	ABR70029	Drosophila melanog	455	6	3.7	365	21	AAB56813	Human prostate can
383	6	3.7	273	23	ABR92422	Herbically activ	456	6	3.7	365	21	AAG06592	Arabidopsis thalia
384	6	3.7	275	21	AAG25357	Arabidopsis thalia	457	6	3.7	365	21	AAG45295	Arabidopsis thalia
385	6	3.7	275	23	ABR32472	Staphylococcus aur	458	6	3.7	365	23	AAU11305	Thermotoga maritim
386	6	3.7	280	21	AAG32935	Arabidopsis thalia	459	6	3.7	366	23	ABP39520	Staphylococcus epi
387	6	3.7	281	18	AAW27560	Consensus single c	460	6	3.7	366	23	AAU10816	Human Type II GnRH
388	6	3.7	281	21	AAG39878	Arabidopsis thalia	461	6	3.7	367	21	AAG39984	Arabidopsis thalia
389	6	3.7	281	22	ABR11315	Human HSP-124 pro	462	6	3.7	367	23	ABR92915	Herbically activ
390	6	3.7	284	23	ABR32484	Staphylococcus aur	463	6	3.7	368	22	AAB82589	Human prothrombina
391	6	3.7	291	21	AAG39985	Arabidopsis thalia	464	6	3.7	368	23	AAU11307	Variant fusion pro
392	6	3.7	291	21	AAB29468	Cucumber xylen sap	465	6	3.7	369	23	AAU11306	Fusion protein fro
393	6	3.7	293	22	ABG06263	Novel human diagno	466	6	3.7	370	22	AAG70708	S cerevisiae apopt
394	6	3.7	295	22	ABG06263	G-protein coupled	467	6	3.7	371	22	AAG83100	S. epidermidis ope
395	6	3.7	297	15	AAK48736	G-protein coupled	468	6	3.7	372	21	AAG46626	Arabidopsis thalia
396	6	3.7	297	15	AAK48737	G-protein coupled	469	6	3.7	372	22	AAE07865	Novel human protei
397	6	3.7	297	17	AAW02708	G-protein coupled	470	6	3.7	372	22	AAG06591	Arabidopsis thalia
398	6	3.7	297	21	AAG30795	Arabidopsis thalia	471	6	3.7	376	21	AAG45294	Arabidopsis thalia
399	6	3.7	297	23	ABP26875	Staphylococcus epi	472	6	3.7	376	23	ABP39414	Scaphylococcus epi
400	6	3.7	297	23	ABR90824	Herbically activ	473	6	3.7	376	23	AAU10815	Human Type II GnRH
401	6	3.7	301	22	ABR63512	Drosophila melanog	474	6	3.7	377	22	ABR61905	Drosophila melanog
402	6	3.7	302	21	AAG39230	Arabidopsis thalia	475	6	3.7	377	23	AAU10814	Human Type II GnRH
403	6	3.7	302	22	AAU40372	Propionibacterium	476	6	3.7	377	23	AAU10819	Human Type II GnRH
404	6	3.7	302	22	AAE02526	Arabidopsis thalia	477	6	3.7	378	22	AAM39027	Human polypeptide
405	6	3.7	305	21	AAG05315	Arabidopsis thalia	478	6	3.7	379	20	AAV25768	Human secreted pro
406	6	3.7	305	21	AAG39980	Arabidopsis thalia	479	6	3.7	379	21	AAB56086	Human secreted pro
407	6	3.7	305	22	AAU58771	Propionibacterium	480	6	3.7	379	23	AAU10813	Human Type II GnRH
408	6	3.7	306	22	AAAB82593	Human prothrombina	481	6	3.7	380	21	AAG25355	Arabidopsis thalia
409	6	3.7	306	23	ABP51325	Human MDRF SEQ ID	482	6	3.7	380	23	AAU10812	Marmoset Type II G
410	6	3.7	307	23	ABP38393	Staphylococcus epi	483	6	3.7	381	21	AAG14473	Arabidopsis thalia
411	6	3.7	308	21	AAG43295	Arabidopsis thalia	484	6	3.7	382	22	ABR76934	Rat VG41. Rattus
412	6	3.7	309	20	AAV37181	Protein involved i	485	6	3.7	383	23	AAG05750	Arabidopsis thalia
413	6	3.7	314	22	ABG10894	Novel human diagno	486	6	3.7	383	23	AAU11304	Human Type II GnRH
414	6	3.7	315	22	ABG11079	Novel human diagno	487	6	3.7	385	21	AAG45293	Arabidopsis thalia
415	6	3.7	315	22	AAK10324	Human transporter	488	6	3.7	386	20	AAV38873	Neisseria meningit
416	6	3.7	315	22	ABR94269	Human protein sequ	489	6	3.7	386	21	AAG06590	Arabidopsis thalia
417	6	3.7	315	22	ABR94664	Human protein sequ	490	6	3.7	386	23	ABR53527	Lactococcus lactis
418	6	3.7	318	12	AAK15480	NP-YA. Homo sapie	491	6	3.7	387	22	AAB95516	Human protein sequ
419	6	3.7	318	21	AAG09036	Arabidopsis thalia	492	6	3.7	388	21	AAG08797	Arabidopsis thalia
420	6	3.7	322	22	AAG92292	C glutamincum prote	493	6	3.7	388	22	ABG07515	Novel human diagno
421	6	3.7	322	22	AAG93157	C glutamincum prote	494	6	3.7	388	22	ABG07515	Novel human diagno
422	6	3.7	323	18	AAM34200	Streptomyces efflu	495	6	3.7	388	22	ABG20786	Novel human diagno
423	6	3.7	323	19	AAW55801	Streptomyces roseo	496	6	3.7	388	22	ABR82584	Human prothrombina
424	6	3.7	326	21	AAG05752	Arabidopsis thalia	497	6	3.7	388	22	AAV72622	Human angiotensin
425	6	3.7	326	22	AAU14275	Human novel protei	498	6	3.7	388	23	ABR77546	Neisseria meningit
426	6	3.7	326	22	AAU14511	Human novel protei	499	6	3.7	389	21	AAV74909	Neisseria meningit
427	6	3.7	326	22	AAAB82590	Human prothrombina	500	6	3.7	391	21	AAG46625	Arabidopsis thalia
428	6	3.7	331	22	AAU50916	Propionibacterium	501	6	3.7	391	21	AAV70005	Casein kinase II a
429	6	3.7	331	22	AAG89937	C glutamincum prote	502	6	3.7	394	21	AAG05314	Arabidopsis thalia
430	6	3.7	335	21	AAB56997	Human prostate can	503	6	3.7	394	21	AAG39979	Arabidopsis thalia
431	6	3.7	335	22	ABR66970	Drosophila melanog	504	6	3.7	394	23	ABR90861	Herbically activ
432	6	3.7	335	22	AAG90139	C glutamincum prote	505	6	3.7	395	22	AAE08013	Rat chimeric rNPY5
433	6	3.7	344	23	AAR24030	Human HPR2 spliced	506	6	3.7	398	21	AAG20852	Arabidopsis thalia
434	6	3.7	345	22	ABR63738	Drosophila melanog	507	6	3.7	398	21	AAK43104	Arabidopsis thalia
435	6	3.7	346	22	AAG65608	Human oocyte speci	508	6	3.7	398	22	ABR59789	Arabidopsis thalia
436	6	3.7	347	22	AAU41440	Propionibacterium	509	6	3.7	401	21	AAG12960	Arabidopsis thalia
437	6	3.7	348	22	ABR04964	Human adenovirus t	510	6	3.7	402	21	AAG39978	Arabidopsis thalia
438	6	3.7	349	22	ABR96784	Putative P. abyssi	511	6	3.7	403	21	AAG05313	Arabidopsis thalia
439	6	3.7	350	22	AAK48107	Arabidopsis thalia	512	6	3.7	403	21	AAK14472	Arabidopsis thalia
440	6	3.7	350	22	ABR59032	Drosophila thalia	513	6	3.7	404	23	ABP27853	Streptococcus poly
441	6	3.7	352	21	AAG32355	Arabidopsis melanog	514	6	3.7	405	20	AAV38875	Neisseria meningit
442	6	3.7	352	22	AAE08011	Arabidopsis thalia	515	6	3.7	405	20	AAV38876	Neisseria gonorrho
443	6	3.7	353	23	AAU79045	Rat chimeric rNPY5	516	6	3.7	405	22	AAG92658	C glutamincum prote
444	6	3.7	354	22	AAU01214	Human novel human	517	6	3.7	405	22	AAAB79146	Corynebacterium gl
445	6	3.7	354	23	AAU011303	Thermotoga maritim	518	6	3.7	406	22	AAAB93320	Human protein sequ
446	6	3.7	355	22	ABG03603	Novel human diagno	519	6	3.7	406	22	AAB68523	Human GTP-binding
447	6	3.7	356	21	AAG05751	Arabidopsis thalia	520	6	3.7	407	20	AAV38874	Neisseria meningit
448	6	3.7	358	21	AAB16412	Pinus radiata diph	521	6	3.7	407	21	AAG08796	Arabidopsis thalia

668	6	3.7	616	19	AAW55977	Aspergillus awamori	741	6	3.7	838	23	ABR08523	Amino acid sequenc
669	6	3.7	616	19	AAW55979	Aspergillus awamori	742	6	3.7	839	13	AAK22275	Bovine liver GPI-P
670	6	3.7	616	19	AAW55980	Aspergillus awamori	743	6	3.7	839	16	AAK75109	Glycosyl-phosphati
671	6	3.7	616	21	AAAB15176	Aspergillus awamori	744	6	3.7	845	22	AAE07868	Novel human protei
672	6	3.7	616	21	AAAB15178	Aspergillus awamori	745	6	3.7	873	22	AAK82971	S. epidermidis ope
673	6	3.7	616	21	AAAB15179	Aspergillus awamori	746	6	3.7	895	22	ABK63617	Drosophila melanog
674	6	3.7	616	21	AAAB15180	Aspergillus awamori	747	6	3.7	897	23	ABK38755	Staphylococcus epi
675	6	3.7	616	21	AAAB15181	Aspergillus awamori	748	6	3.7	900	23	ABK3716	Herbicidally activ
676	6	3.7	616	21	AAAB15184	Aspergillus awamori	749	6	3.7	921	15	AAK56487	TATP-binding prote
677	6	3.7	616	21	AAAG6707	Arabidopsis thaliana	750	6	3.7	921	17	AAW06077	Drosophila TATA-bi
678	6	3.7	621	19	AAW55978	Aspergillus awamori	751	6	3.7	921	18	AAW25028	TATP-binding prote
679	6	3.7	621	21	AAAB15182	Aspergillus awamori	752	6	3.7	921	22	ABK61528	Drosophila melanog
680	6	3.7	621	21	AAAB15183	Aspergillus awamori	753	6	3.7	921	22	ABK66055	Drosophila melanog
681	6	3.7	621	23	AAU11270	Human coagulation	754	6	3.7	924	22	ABK30172	Novel human diagra
682	6	3.7	621	23	AAU11271	Human coagulation	755	6	3.7	924	22	AAK33084	Human protein sequ
683	6	3.7	630	9	AAK81876	Sequence of amylog	756	6	3.7	924	22	AAK35845	Human protein sequ
684	6	3.7	631	22	ABG22173	Novel human diagra	757	6	3.7	925	21	AAK30684	Arabidopsis thaliana
685	6	3.7	632	16	AAK71373	Human GRK #1. Hom	758	6	3.7	925	23	ABK33251	Herbicidally activ
686	6	3.7	637	15	AAK63673	Glutamylcysteine-s	759	6	3.7	930	19	AAW57260	Human semaphorin Y
687	6	3.7	639	5	AAK40212	Sequence encoded b	760	6	3.7	930	20	AAK55052	Chlamydia pneumoni
688	6	3.7	640	21	AAK77741	A. niger G1 glucosa	761	6	3.7	930	20	AAK88424	Chlamydia pneumoni
689	6	3.7	640	22	AAK61905	A. niger protein s	762	6	3.7	930	21	AAK90240	Chlamydia anligen
690	6	3.7	650	22	AAE07867	Novel human protei	763	6	3.7	930	21	ABK90548	Chlamydia pneumoni
691	6	3.7	655	23	AAU76822	Rice clone rdl1c-P	764	6	3.7	936	21	AAK30683	Arabidopsis thaliana
692	6	3.7	658	21	AAK31406	Novel human diagra	765	6	3.7	950	21	AAK53899	Amino acid sequenc
693	6	3.7	663	22	ABK08738	Novel human diagra	766	6	3.7	958	22	ABK06858	Novel human diagra
694	6	3.7	669	22	ABK61302	Drosophila melanog	767	6	3.7	968	22	ABK68484	Drosophila melanog
695	6	3.7	669	22	ABK47564	Protease PR7S-6.	768	6	3.7	973	18	AAK41815	A. thaliana trehal
696	6	3.7	669	23	AAU79044	Human novel human	769	6	3.7	973	18	AAK41045	Human polypeptide
697	6	3.7	669	23	AAU79046	Human novel human	770	6	3.7	975	22	ABK24301	Novel human diagra
698	6	3.7	669	23	AAU72883	Human aspartyl pro	771	6	3.7	984	22	ABK63604	Drosophila melanog
699	6	3.7	675	18	AAK06449	Brassic napus FCA	772	6	3.7	994	22	AAK44844	S. lepidophylla tr
700	6	3.7	679	21	AAK58294	Lung cancer associ	773	6	3.7	997	22	ABK59064	Novel human diagra
701	6	3.7	684	20	AAK25114	Mouse alpha1 (XVII	774	6	3.7	1027	18	AAK17057	Drosophila melanog
702	6	3.7	691	21	AAK31405	Arabidopsis thaliana	775	6	3.7	1054	22	ABK57993	Candida albicans c
703	6	3.7	691	21	AAK51807	Arabidopsis thaliana	776	6	3.7	1057	23	ABK2982	Drosophila melanog
704	6	3.7	692	22	AAK65619	Novel protein kina	777	6	3.7	1058	22	ABK71401	Herbicidally activ
705	6	3.7	695	22	ABK59756	Drosophila melanog	778	6	3.7	1081	22	ABK66277	Drosophila melanog
706	6	3.7	701	21	AAK31685	Arabidopsis thaliana	779	6	3.7	1081	22	ABK71824	Drosophila melanog
707	6	3.7	701	22	AAK92442	C glutamincum prote	780	6	3.7	1081	22	ABK71824	Phototaxis lumini
708	6	3.7	720	21	AAK51806	Novel human protei	781	6	3.7	1095	18	AAK17868	Toxin TcaA, encode
709	6	3.7	724	22	AAE07866	Novel human protei	782	6	3.7	1095	19	AAK56547	Novel human diagra
710	6	3.7	727	21	AAK31404	Arabidopsis thaliana	783	6	3.7	1116	22	ABK22811	Herbicidally activ
711	6	3.7	727	23	ABK3080	Herbicidally activ	784	6	3.7	1134	23	ABK3262	Drosophila melanog
712	6	3.7	730	21	AAK31684	Arabidopsis thaliana	785	6	3.7	1164	22	ABK61495	Drosophila melanog
713	6	3.7	734	22	ABK63238	Drosophila melanog	786	6	3.7	1166	20	AAK08643	S. aureus SdrE pro
714	6	3.7	734	22	ABK63239	Drosophila melanog	787	6	3.7	1174	22	ABK65304	Drosophila melanog
715	6	3.7	747	22	AAE14444	Drosophila melanog	788	6	3.7	1219	22	ABK65388	Drosophila melanog
716	6	3.7	751	22	AAE14444	Human drug metabol	789	6	3.7	1225	19	AAK52289	Homo sapiens cdo t
717	6	3.7	759	22	ABK07769	Novel human diagra	790	6	3.7	1226	21	AAK42239	Human polypeptide
718	6	3.7	759	22	ABK08475	Novel human diagra	791	6	3.7	1278	22	AAK39259	Human polypeptide
719	6	3.7	760	21	AAK51805	Arabidopsis thaliana	792	6	3.7	1288	18	AAK6338	Mouse alpha-1 (XVI
720	6	3.7	766	23	AAE21719	Human PKIN-14 prot	793	6	3.7	1288	20	AAK22927	Drosophila melanog
721	6	3.7	766	23	AAE21719	Human serine/threos	794	6	3.7	1297	22	ABK64155	Drosophila melanog
722	6	3.7	766	23	ABK92836	Herbicidally activ	795	6	3.7	1299	22	ABK70025	Drosophila melanog
723	6	3.7	770	20	AAK13456	Amino acid sequenc	796	6	3.7	1310	22	ABK60913	Arabidopsis thaliana
724	6	3.7	770	21	AAK31683	Arabidopsis thaliana	797	6	3.7	1321	21	AAK42238	Arabidopsis thaliana
725	6	3.7	771	22	ABK65630	Drosophila melanog	798	6	3.7	1351	21	AAK42237	Arabidopsis thaliana
726	6	3.7	771	22	ABK07869	Novel human protei	799	6	3.7	1352	21	AAK31776	Arabidopsis thaliana
727	6	3.7	785	22	AAK30090	C glutamincum prote	800	6	3.7	1360	22	AAU11802	Human protein kina
728	6	3.7	791	22	AAK39420	Human polypeptide	801	6	3.7	1374	21	AAK11934	Human MEKK5, Homo
729	6	3.7	793	22	ABK71688	Drosophila melanog	802	6	3.7	1375	18	AAK27283	Apoptosis inducing
730	6	3.7	796	21	AAK30685	Arabidopsis thaliana	803	6	3.7	1377	23	ABK77522	Ring finger protei
731	6	3.7	802	22	AAK47592	Human DHDR-1. Hom	804	6	3.7	1400	20	AAK82791	Human RON receptor
732	6	3.7	803	22	AAU87246	Novel central nerv	805	6	3.7	1404	23	AAK20551	Proliferation pote
733	6	3.7	806	22	ABK21109	Novel human diagra	806	6	3.7	1406	21	AAK31775	Arabidopsis thaliana
734	6	3.7	808	20	AAK49873	Herbicidally ionotr	807	6	3.7	1407	22	AAK31774	Arabidopsis thaliana
735	6	3.7	808	22	ABK92200	Herbicidally activ	808	6	3.7	1407	22	AAK86123	A. thaliana formyl
736	6	3.7	809	22	ABK59391	Drosophila melanog	809	6	3.7	1446	22	ABK3929	Drosophila melanog
737	6	3.7	809	22	ABK91284	Herbicidally activ	810	6	3.7	1462	23	ABK7521	Novel human protei
738	6	3.7	827	18	AAK07686	Ionotropic glutama	811	6	3.7	1499	22	ABK58409	Drosophila melanog
739	6	3.7	831	22	AAK91714	C glutamincum prote	812	6	3.7	1508	22	AAK50676	C. elegans alpha-2
740	6	3.7	837	22	AAK41206	Human polypeptide	813	6	3.7	1519	22	AAK50677	C. elegans alpha-2

814	6	3.7	1528	17	AAR95333	Manduca sexta Bac1	887	3.7	6239	21	AAB23750	S. avermiltitis ave
815	6	3.7	1528	20	AAW90182	Manduca sexta Br t	888	3.7	6239	22	AAAG65265	Streptomyces averm
816	6	3.7	1566	22	ABR64864	Drosophila melanog	889	3.7	6685	22	ABG23030	Novel human diagno
817	6	3.7	1617	22	AAE07870	Novel human protei	890	3.7	7201	22	ABB71136	Drosophila melanog
818	6	3.7	1635	22	ABR90958	Herbicidially activ	891	3.7	10182	22	ABP38314	Staphylococcus epi
819	6	3.7	1653	22	AAAG98354	Escherichia coli p	892	3.1	7	20	AAW92343	Latent fluorescent
820	6	3.7	1691	22	AAE10609	Human novel KIAA12	893	3.1	8	19	AAW57875	C. felis esterase
821	6	3.7	1691	22	AAE07863	Novel human protei	894	3.1	8	22	AAE12921	Ctenocephalides fe
822	6	3.7	1707	22	AAAG78760	Hepatitis E virus	895	3.1	9	21	AAV97234	Complementary dete
823	6	3.7	1752	23	ABR05030	Bifidobacterium bi	896	3.1	9	22	AAE13142	Humanised antibody
824	6	3.7	1697	21	AAAB19712	Human protein tyro	897	3.1	9	22	AAAG79232	Peptide of 45 kDa
825	6	3.7	1697	21	AAV81785	LAR tyrosine phosph	898	3.1	9	22	AAAB82708	VEGF antagonist an
826	6	3.7	1907	22	AAU14143	Human novel protei	899	3.1	9	22	AAAG63993	Mutant scFv 508F-1
827	6	3.7	1907	22	AAU14143	Drosophila melanog	900	3.1	9	22	AAU15039	Complementarily de
828	6	3.7	1916	22	ABR62423	Drosophila melanog	901	3.1	9	22	AAU15040	Antiangiogenic pep
829	6	3.7	1916	22	ABR656089	Drosophila melanog	902	3.1	9	23	AAU95918	Antiangiogenic pep
830	6	3.7	1940	22	ABR64070	Drosophila melanog	903	3.1	9	23	AAU95918	Immunogenic peptid
831	6	3.7	1963	22	ABR58053	Drosophila melanog	904	3.1	9	23	AAU74411	Light chain comple
832	6	3.7	2070	22	ABG03703	Novel human diagno	905	3.1	9	23	AAE17068	Measles virus V pr
833	6	3.7	2088	22	AAAM80124	Human proteoln SEQ	906	3.1	10	18	AAW31847	Releas factor GRF
834	6	3.7	2090	22	ABR64682	Drosophila melanog	907	3.1	10	22	AAAG39902	Human complementar
835	6	3.7	2114	22	ABG09076	Novel human diagno	908	3.1	10	22	AAAG94138	Human complementar
836	6	3.7	2166	22	AAW79140	Human proteoln SEQ	909	3.1	10	22	AAAG97006	Human complementar
837	6	3.7	2209	23	AAU85609	Lung tumour protei	910	3.1	10	22	AAAG84202	Arabidopsis thalia
838	6	3.7	2215	22	AAAG64030	Human polypeptide	911	3.1	10	22	AAAG85675	Saccharomyces cere
839	6	3.7	2356	22	ABG09072	Novel human diagno	912	3.1	10	22	AAAG85675	Saccharomyces cere
840	6	3.7	2408	13	AAAR24306	Translation of ORF	913	3.1	10	22	AAAG85675	Hsp-65 peptide epi
841	6	3.7	2496	18	AAW36093	Mutant YR087c pro	914	3.1	10	22	AAAG85675	Human KRPI tryptic
842	6	3.7	2713	18	AAW19654	ATM mutant C8140T.	915	3.1	10	22	AAAG85675	Breast cancer-asso
843	6	3.7	2759	18	AAW19652	ATM mutant C8140T.	916	3.1	10	22	AAAG85675	Anti-A3 antigen I
844	6	3.7	2766	18	AAW19651	ATM mutant 8269del	917	3.1	11	21	AAV66797	A3 antigen bindin
845	6	3.7	2768	18	AAW19650	ATM mutant 8283del	918	3.1	11	22	AAV66797	Human ADBI tryptic
846	6	3.7	2932	18	AAW19675	ATM mutant G8307A.	919	3.1	11	22	AAV66797	C. felis esterase
847	6	3.7	2942	18	AAW23050	ATM mutant 2467del	920	3.1	11	22	AAV66797	C. felis esterase
848	6	3.7	2987	18	AAW19649	Saccharomyces cere	921	3.1	11	22	AAV66797	Human secreted pro
849	6	3.7	2987	18	AAW19649	ATM mutant 8946ins	922	3.1	12	19	AAW57871	Ctenocephalides fe
850	6	3.7	2998	18	AAW19688	ATM mutant 1407del	923	3.1	12	19	AAW57871	Human Breast cance
851	6	3.7	2998	18	AAW19691	ATM mutant 3403del	924	3.1	12	22	AAW57871	Human Breast cance
852	6	3.7	3001	18	AAW19673	ATM mutant 4612del	925	3.1	12	22	AAW57871	Colostrinlin deriv
853	6	3.7	3001	18	AAW19670	ATM mutant 4612del	926	3.1	12	22	AAW57871	Colostrinlin deriv
854	6	3.7	3003	18	AAW19658	ATM mutant 4612del	927	3.1	12	22	AAW57871	Colostrinlin deriv
855	6	3.7	3005	18	AAW19653	ATM mutant 7630del	928	3.1	12	22	AAW57871	Colostrinlin deriv
856	6	3.7	3005	18	AAW19648	ATM mutant 8269del	929	3.1	12	22	AAW57871	Colostrinlin deriv
857	6	3.7	3021	18	AAW19661	ATM mutant 9001del	930	3.1	12	22	AAW57871	Colostrinlin deriv
858	6	3.7	3026	18	AAW19676	ATM mutant 6348del	931	3.1	12	22	AAW57871	Colostrinlin deriv
859	6	3.7	3046	18	AAW19690	ATM mutant 2377del	932	3.1	12	22	AAW57871	Colostrinlin deriv
860	6	3.7	3053	18	AAW19694	ATM mutant C9140T.	933	3.1	12	22	AAW57871	Colostrinlin deriv
861	6	3.7	3054	18	AAW19702	ATM mutant 7636del	934	3.1	12	22	AAW57871	Colostrinlin deriv
862	6	3.7	3054	18	AAW19695	ATM mutant 7278del	935	3.1	12	22	AAW57871	Colostrinlin deriv
863	6	3.7	3055	18	AAW19697	ATM mutant 7279del	936	3.1	12	22	AAW57871	Colostrinlin deriv
864	6	3.7	3055	18	AAW19693	ATM mutant 5435del	937	3.1	12	22	AAW57871	Colostrinlin deriv
865	6	3.7	3056	18	AAW06234	ATM mutant 8578del	938	3.1	12	22	AAW57871	Colostrinlin deriv
866	6	3.7	3056	18	AAW06234	ATM mutant 8578del	939	3.1	12	22	AAW57871	Colostrinlin deriv
867	6	3.7	3056	18	AAW06234	ATM mutant 8578del	940	3.1	12	22	AAW57871	Colostrinlin deriv
868	6	3.7	3056	18	AAW19733	Ataxia-telangiecta	941	3.1	12	22	AAW57871	Colostrinlin deriv
869	6	3.7	3056	18	AAW19698	Cell cycle checkpo	942	3.1	12	22	AAW57871	Colostrinlin deriv
870	6	3.7	3056	19	AAW37133	ATM mutant E2904G.	943	3.1	12	22	AAW57871	Colostrinlin deriv
871	6	3.7	3056	20	AAW84269	Ataxia-telangiecta	944	3.1	12	22	AAW57871	Colostrinlin deriv
872	6	3.7	3056	20	AAW84269	Ataxia-telangiecta	945	3.1	12	22	AAW57871	Colostrinlin deriv
873	6	3.7	3056	22	AAAB62280	Human ataxia-telan	946	3.1	12	22	AAW57871	Colostrinlin deriv
874	6	3.7	3056	22	AAAB62280	Human ataxia-telan	947	3.1	12	22	AAW57871	Colostrinlin deriv
875	6	3.7	3056	22	AAAB62280	Human ataxia-telan	948	3.1	12	22	AAW57871	Colostrinlin deriv
876	6	3.7	3056	22	AAAB62280	Human ataxia-telan	949	3.1	12	22	AAW57871	Colostrinlin deriv
877	6	3.7	3057	22	AAAB68896	Human ATM proteoln	950	3.1	12	22	AAW57871	Colostrinlin deriv
878	6	3.7	3059	18	AAW19696	ATM mutant 5319ins	951	3.1	12	22	AAW57871	Colostrinlin deriv
879	6	3.7	3066	18	AAW36178	Marine Ataxia-tela	952	3.1	12	22	AAW57871	Colostrinlin deriv
880	6	3.7	3085	18	AAW19701	ATM mutant G9170C.	953	3.1	12	22	AAW57871	Colostrinlin deriv
881	6	3.7	3313	22	AAU30134	Novel human secret	954	3.1	12	22	AAW57871	Colostrinlin deriv
882	6	3.7	4630	21	AAW19629	Streptomyces venez	955	3.1	12	22	AAW57871	Colostrinlin deriv
883	6	3.7	4630	21	AAW19629	Streptomyces venez	956	3.1	12	22	AAW57871	Colostrinlin deriv
884	6	3.7	4725	22	ABG23837	S. venezuelae vep	957	3.1	12	22	AAW57871	Colostrinlin deriv
885	6	3.7	4977	22	ABG17057	Novel human diagno	958	3.1	12	22	AAW57871	Colostrinlin deriv
886	6	3.7	6025	22	AAAG84939	Novel human diagno	959	3.1	12	22	AAW57871	Colostrinlin deriv
						Shrimp white spot						

960	5	3.1	14	22	AAB59353	Ewe colostrin pe
961	5	3.1	15	17	AAP94846	peptide from libra
962	5	3.1	15	17	AAR94847	peptide from libra
963	5	3.1	15	18	AAW43525	Mycobacteria sp. h
964	5	3.1	15	18	AAW43524	Mycobacteria sp. h
965	5	3.1	15	18	AAW23522	Purified cis-9,10-
966	5	3.1	15	19	AAW20515	Human neurofilamen
967	5	3.1	15	19	AAW57801	Fatty acid amide h
968	5	3.1	15	19	AAW62158	Agrobacterium faec
969	5	3.1	15	19	AAW62159	Agrobacterium faec
970	5	3.1	15	19	AAW62167	Agrobacterium faec
971	5	3.1	15	20	AAW43168	Integrin cell surf
972	5	3.1	15	21	AAW29733	Ganglioside GM1-bi
973	5	3.1	15	21	AAW92952	Transforming growt
974	5	3.1	15	21	AAW93104	Transforming growt
975	5	3.1	15	21	AAW66922	T cell antigen rec
976	5	3.1	15	21	AAW51055	EcotinAPolI derived
977	5	3.1	15	21	AAW54488	Estradiol activat
978	5	3.1	15	22	AAW88294	Hsp-65 peptide epi
979	5	3.1	15	23	ABW09948	Integrin beta.6 su
980	5	3.1	15	23	AAO18022	Human immunoglobul
981	5	3.1	15	23	ABU00523	Human 19E cyclic l
982	5	3.1	15	23	AAW086212	Oestradiol activat
983	5	3.1	15	23	ABW06199	Human DNA methylat
984	5	3.1	15	23	ABW03468	Vascular endothell
985	5	3.1	15	23	ABW04314	Human Pax protein
986	5	3.1	16	21	AAW11423	Coxsackievirus B C
987	5	3.1	16	22	AAW88293	Hsp-65 peptide epi
988	5	3.1	16	23	AAW24517	Human RAR1d6 N-my
989	5	3.1	16	23	AAW67867	Human ADP1 tryptic
990	5	3.1	16	23	AAW17067	Measles virus V pr
991	5	3.1	17	17	AAW01973	Rat pro-opiomelano
992	5	3.1	18	21	AAW22069	p67SRF protein fra
993	5	3.1	18	21	AAW22069	p67SRF protein fra
994	5	3.1	18	21	AAW5867	O-glycosylation s1
995	5	3.1	18	22	AAW99477	p67-SRF peptide #3
996	5	3.1	18	22	AAW88292	Hsp-65 peptide epi
997	5	3.1	18	23	AAW75512	Transcription fact
998	5	3.1	19	17	AAW94848	Peptide from libra
999	5	3.1	19	18	AAW43526	Mycobacteria sp. h
1000	5	3.1	19	21	AAW51540	Yada homologous pe

ALIGNMENTS

RESULT 1
AAE18419 standard; Protein: 279 AA.

XX	AAE18419;	
AC		
DT	07-MAY-2002 (first entry)	
XX		
DE	Escherichia coli strain B240 FimH protein.	
XX		
KW	FimH; immune response; antibacterial; enterobacillus-related disease;	
KW	therapy; vaccine; urinary tract infection; bladder.	
XX		
OS	Escherichia coli B240.	
XX		
FN	Key	Location/Qualifiers
FT	Misc-difference 201	
FT	/note- "Encoded by ACC"	
XX		
FN	WO200204496-A2.	
XX		
PD	17-JAN-2002.	
XX		
PE	06-JUL-2001; 2001WO-US21525.	
XX		
PR	07-JUL-2000; 2000US-216750P.	
XX		

PA	(MED1-) MEDIMMUNE INC.
XX	
PI	Langermann S., Revel A., Auguste C., Burtlein J;
XX	
DR	WPI; 2002-171702/22.
DR	N-PSDB; AAD29358.
XX	
PT	New immunogenic polypeptide, useful as vaccine for protecting against
PT	an enterobacillus-related disease in a patient at risk of contracting
PT	such disease, e.g. urinary tract infection or a bladder infection
XX	
PS	Claim 3; Fig 2; 101pp; English.
XX	
CC	The invention relates to bacterial immunogenic agents for administration
CC	to humans and non-human animals to stimulate an immune response. The
CC	invention also relates to methods for vaccination of mammalian species
CC	with variants of E. coli FimH protein derived from different strains of
CC	E. coli. The vaccine composition or the antibody is useful for protecting
CC	against and treating an enterobacillus-related disease in a patient
CC	afflicted or at a risk of contracting the disease. In particular, the
CC	disease is a urinary tract or bladder infection. The disease is caused
CC	by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC	The present sequence is Escherichia coli strain B240 FimH protein.
XX	
SQ	Sequence 279 AA;

Query Match 100.0%; Score 161; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.4e-156;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1 PVVVVGQNLVVDLSTQIFCHNDYPETITDVTYTLQSGAYGVLISNFGYVYSGSSYPFP 60
DB	26 PVVVVGQNLVVDLSTQIFCHNDYPETITDVTYTLQSGAYGVLISNFGYVYSGSSYPFP 85
QY	61 TTSETPRVYVNSRDKPMPVALYLTLPVSSAGGLYIKGSLIAVLILROTNNYNSDDFOFV 120
DB	86 TTSETPRVYVNSRDKPMPVALYLTLPVSSAGGLYIKGSLIAVLILROTNNYNSDDFOFV 145
QY	121 WNIYANNDDVVPPTGCDVSARDVVTLPDVRGSVPILPTVY 161
DB	146 WNIYANNDDVVPPTGCDVSARDVVTLPDVRGSVPILPTVY 186

RESULT 2
ABG68154 standard; Protein: 268 AA.

XX	ABG68154;	
AC		
DT	07-OCT-2002 (first entry)	
XX		
DE	Escherichia coli major fimbrial protein FimH.	
XX		
KW	Immunogen; FimH adhesin functional site-derived target peptide;	
KW	FARSD; helper T cell epitope; FimH; urinary tract infection;	
KW	type 1 fimbriated uropathogenic enterobacteria; vaccine;	
KW	FARSD site-specific immunity.	
XX		
OS	Escherichia coli.	
XX		
FN	WO200251860-A2.	
XX		
PD	04-JUL-2002.	
XX		
PE	21-DEC-2001; 2001WO-US50816.	
XX		
PR	22-DEC-2000; 2000US-0747802.	
XX		
PA	(UNBI-) UNITED BIOMEDICAL INC.	
XX		
PI	Wang CY;	
XX		
DR	WPI; 2002-528681/56.	

XX Novel peptide immunogen, useful for evoking antibodies to prevent
PT adherence of *Escherichia coli* to bladder mucosa, comprises a FimH
PT adhesin functional site-derived target peptide covalently linked to
PT helper T cell epitope -
XX
PS Disclosure; Fig 1; 62pp; English.
XX
CC The invention describes a peptide immunogen (I), comprising a helper T
CC cell epitope sequence (T_H) or a carrier protein covalently attached to a
CC FimH adhesin functional site-derived (FAFSD) target peptide comprising
CC not more than 30 amino acids of the carbohydrate binding pocket of FimH,
CC or its crossreactive and immunologically functional analogue or mimotope.
CC (I) and a composition containing (I) are useful for inducing anti-FAFSD
CC peptide antibody production in a mammal. The composition is also useful
CC for reducing adherence to the urinary tract mucosa of a mammal by type 1
CC fimbriated uropathogenic enterobacteria (*Escherichia coli*) to prevent
CC urinary tract infection. (I) has a focused FAFSD site-specific immunity
CC together with a broad protective immunity, and with less adverse side
CC reactions than the more complex polypeptide subunit vaccines and the
CC carrier conjugated vaccine. Since (I) is chemically well defined it is
CC easy and less costly to manufacture and to control or assure the quality
CC of the product. This is the amino acid sequence of FimH adhesin, peptides
CC derived from the functional site are used in the creation of a vaccine
CC against urinary tract infection described in the invention.
XX
SQ Sequence 268 AA:

Query Match 57.1%; Score 92; DB 23; Length 268;
Best Local Similarity 100.0%; Pred. No. 7.1e-86;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PNVNQGNIIVLDISTQIFCHNDYPERITDVTYLRGSAYGVLNFGTVKYGSSSYPP 60
Db 26 PNVNQGNIIVLDISTQIFCHNDYPERITDVTYLRGSAYGVLNFGTVKYGSSSYPP 85
QY 61 TTSETPRVYVNSRTDKPMPVALYLRPVSSAG 92
Db 86 TTSETPRVYVNSRTDKPMPVALYLRPVSSAG 117

RESULT 3
AAB47074
ID AAB47074 standard; Protein; 279 AA.
XX
AC AAB47074;
XX
DT 08-MAY-2001 (first entry)
XX
DE Adhesin protein, FimH.
XX
KW Donor: pilus protein; pilin; adhesin; vaccine; urinary epithelia;
KW urinary tract infection; enterobacteriaceae.
XX
OS *Escherichia coli*.
XX
PN WO200104148-A2.
PD 18-JAN-2001.
PF 13-JUL-2000; 2000WO-US19066.
PR 13-JUL-1999; 99US-0143582.
PR 16-JUL-1999; 99US-0144359.
PR 23-FEB-2000; 2000US-0184442.
XX
PA (MEDT-) MEDIMMUNE INC.
PI Hultgren SJ, Pinkner JS, Sauer F, Barnhart M, Waksman G, Knight S;
XX
DR WPI; 2001-138315/14.
XX
CC Immunogenic complexes and polypeptides for vaccinating against urinary

PT tract disease, comprises a pilus protein component and a bacterial
PT chaperrone -
XX
PS Disclosure; Fig 1; 92pp; English.
XX
CC This sequence may be used as the pilus protein in the immunogenic
CC complex of the invention. The complex comprises a pilus protein
CC component and a donor complement portion as part of the same amino
CC acid sequence or as non-covalently linked fragments of a complex
CC such that the correct conformation of the pilin is maintained. The
CC pilus protein component may be an adhesin or a pilin. Pilus associated
CC adhesins, such as FimH are relatively conserved proteins among
CC different species and strains of bacteria, therefore vaccines
CC incorporating the FimH antigen exhibit a broad spectrum of
CC protection compared with current pilus-fiber based vaccines. The
CC immunogenic complexes act by disrupting pilus-mediated attachment
CC of *E. coli* to urinary epithelia and may prevent or retard the
CC development of urinary tract infections. Vaccines containing the
CC complexes are useful for preventing urinary tract disease in a human
CC caused by the bacterium family enterobacteriaceae specifically
CC *Escherichia coli* and may also be used in treating the disease.
XX
SQ Sequence 279 AA:

Query Match 57.1%; Score 92; DB 22; Length 279;
Best Local Similarity 100.0%; Pred. No. 7.4e-86;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PNVNQGNIIVLDISTQIFCHNDYPERITDVTYLRGSAYGVLNFGTVKYGSSSYPP 60
Db 26 PNVNQGNIIVLDISTQIFCHNDYPERITDVTYLRGSAYGVLNFGTVKYGSSSYPP 85
QY 61 TTSETPRVYVNSRTDKPMPVALYLRPVSSAG 92
Db 86 TTSETPRVYVNSRTDKPMPVALYLRPVSSAG 117

RESULT 4
AAB72839
ID AAB72839 standard; Protein; 279 AA.
XX
AC AAB72839;
XX
DT 09-MAY-2001 (first entry)
XX
DE Bacterial type 1 pilus subunit FimA.
XX
KW Antibacterial compound; Gram-negative bacterium; pilus; chaperrone;
KW biofilm; disease treatment; bacterial infection.
XX
OS Bacteria.
XX
PN WO200110386-A2.
PD 15-FEB-2001.
PF 11-AUG-2000; 2000WO-US22087.
PR 11-AUG-1999; 99US-0148280.
XX
PA (UNITW) UNIV WASHINGTON.
PI Hultgren SJ, Sauer FG, Waksman G, Fuetterer K;
XX
DR WPI; 2001-226496/23.
XX
XX
XX An isolated compound for inhibiting pilus assembly -
XX
XX Example 5; Fig 8A; 14pp; English.
XX
CC The present invention provides antibacterial compounds which are able to
CC interfere with Gram-negative bacteria pilus formation and assembly, and
CC pilus interaction with chaperrone proteins. These are useful in the

CC treatment of bacterial infection, and in the prevention of biofilm
CC formation. They are particularly useful against *Escherichia coli*,
CC *Haemophilus influenzae*, *Salmonella enteritidis*, *S. typhimurium*,
CC *Bordetella pertussis*, *Yersinia enterocolitica*, *Helicobacter pylori* and
CC *Klebsiella pneumoniae*.

SO Sequence 279 AA;

Query Match 57.18; Score 92; DB 22; Length 279;
Best Local Similarity 100.0%; Pred. No. 7,4e-86;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PYNVNGMLVVDLSTQICNDYPTITDYTLORGSAVGGLNSGTVKSSGSSYPP 60
|||||
Db 26 PYNVNGMLVVDLSTQICNDYPTITDYTLORGSAVGGLNSGTVKSSGSSYPP 85
|||||

Oy 61 TTSETPRVYNSRDKPMPVALYLTPVSSAGS 92
|||||
Db 86 TTSETPRVYNSRDKPMPVALYLTPVSSAGS 117
|||||

RESULT 5
AA72515

ID AAY72515 standard; Protein: 279 AA.

XX AAY72515;

DT 02-MAY-2001 (first entry)

XX *Escherichia coli* FliH protein.

KW FliH; adhesin protein; type 1 pilus; mannose binding domain; MBD; COL;
KW collagen binding domain; prophylaxis; therapy; urinary tract infection;
KW Uti; immunogen; passive immunotherapy; vaccine; antibacterial.

XX *Escherichia coli*.

OS location/Qualifiers

FT Key 1..156
FT Domain /label= Lectin_binding-domain
FT 1..20
FT Domain /label= MBD-1
FT /note= "Mannose binding domain-1"

FT Region 2..5
FT /note= "Beta strand 1"

FT Region 9..11
FT /note= "Beta strand 2"

FT Region 17..25
FT /note= "Beta strand 3"

FT Binding-site 27
FT /label= FimC_chaperone_binding_site

FT Binding-site 31
FT /label= FimC_chaperone_binding_site

FT Region 32..37
FT /note= "Beta strand 4a"

FT Region 38..41
FT /note= "3-10 helix"

FT Region 42..46
FT /note= "Beta strand 4b"

FT Domain 46..54
FT /label= MBD-2

FT Domain 50..80
FT /note= "Mannose binding domain 2"

FT Domain 50..80
FT /label= COL

FT /note= "Collagen binding domain; Forms a beta-sheet-
FT alpha-1-beta-sheet structure from strands 5, alpha-1
FT and 6"

FT Region 54..63
FT /note= "Beta strand 5"

FT Region 67..69
FT /note= "Alpha-1 helix"

FT Region 71..77
FT /note= "Beta strand 6"

FT Region

FT /note= "Beta strand 7"

FT Region 89..95
FT /note= "Beta strand 8"

FT Region 104..111
FT /note= "Beta strand 9"

FT Binding-site 117
FT /label= FimC_chaperone_binding_site

FT Region 125..135
FT /note= "Beta strand 10"

FT Domain 127..148
FT /label= MBD-3

FT Region 139..150
FT /note= "Mannose binding domain 3"

FT Binding-site 155
FT /label= FimC_chaperone_binding_site

FT Binding-site 157..158
FT /label= FimC_chaperone_binding_site

FT Domain 159..279
FT /label= Pilin_domain

FT Binding-site 160
FT /label= FimC_chaperone_binding_site

FT Region 161..165
FT /note= "Beta strand A'"

FT Binding-site 162..163
FT /label= FimC_chaperone_binding_site

FT Binding-site 168..171
FT /note= "FimC_chaperone binding site; Serves as beta
strand A'"

FT Binding-site 173
FT /label= FimC_chaperone_binding_site

FT Binding-site 181
FT /label= FimC_chaperone_binding_site

FT Binding-site 183
FT /label= FimC_chaperone_binding_site

FT Region 183..187
FT /note= "Beta strand B"

FT Binding-site 191..192
FT /label= FimC_chaperone_binding_site

FT Region 192..197
FT /note= "Beta strand C"

FT Binding-site 196
FT /label= FimC_chaperone_binding_site

FT Binding-site 198
FT /label= FimC_chaperone_binding_site

FT Binding-site 215
FT /label= FimC_chaperone_binding_site

FT Region 221..226
FT /note= "Beta strand D'"

FT Region 235..239
FT /note= "Beta strand D'"

FT Region 251..258
FT /note= "Beta strand E"

FT Binding-site 266..279
FT /label= FimC_chaperone_binding_site

FT Region 268..279
FT /note= "Beta strand F; contains conserved beta-
zipper motif"

WO200105978-A1.

PD 25-JAN-2001.

PF 14-JUL-2000; 2000WO-US19402.

PR 15-JUL-1999; 99US-0144016.

PA (MEDI-) MEDIMUNE INC.

PI Hultgren SJ, Langermann S;

DR WP1; 2001-159539/16.

XX Polypeptides useful as vaccines for prevention and/or treatment of
PT diseases such as urinary tract infections, caused by
PT Enterobacteriaceae, comprises mannose-binding domains derived from
XX adhesin molecules -
XX
PS Claim 1; Fig 3; 53pp; English.
XX
CC The present sequence is FliH protein from Escherichia coli. FliH
CC is an adhesin protein found in type 1 pili of bacteria of the
CC family enterobacteriaceae, especially E. coli. The FliH
CC protein comprises mannose-binding domains (MBDs) and collagen-
CC binding domains (COL). The present invention relates to
CC engineered polypeptides comprising one or more domains derived
CC from FliH protein. These polypeptides are used to produce
CC prophylactic vaccines which are useful for the prevention
CC and/or treatment of diseases, such as urinary tract infection (UTI)
CC caused by a bacterium of the family Enterobacteriaceae, especially
CC E. coli in animals, in particular humans. They are useful as immunogens
CC to stimulate the production of antibodies for use in passive immuno-
CC therapy, as a diagnostic reagent and as a reagent in other processes
CC such as affinity chromatography. The antibodies of the novel poly-
CC peptides are also useful for research purposes for studying protein-
CC lectin or collagen binding interactions.
XX
SQ Sequence 279 AA:
Query Match 57.1%; Score 92; DB 22; Length 279;
Best Local Similarity 100.0%; Pred. No. 7.4e-86;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PNVNNGNLVVDLSTQIFCHNDYPETITDVTYLRGSAVGVLSNFGTVKYSGSSYPP 60
DB 26 PNVNNGNLVVDLSTQIFCHNDYPETITDVTYLRGSAVGVLSNFGTVKYSGSSYPP 85
QY 61 TTSETPRVYVNSRTDKPMPVALYLRPVSSAGG 92
DB 86 TTSETPRVYVNSRTDKPMPVALYLRPVSSAGG 117
RESULT 6
AAE18424
ID AAE18424 standard; Protein; 279 AA.
XX
AC AAE18424;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC56 FliH protein.
XX
KM FliH: Immune response; antibacterial; enterobacillus-related disease;
KM therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli EC56.
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlain J;
XX
DR MPI; 2002-171702/22.
DR N-PSDB; AAD29363.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -

XX
PS Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FliH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
XX
SQ Sequence 279 AA:
Query Match 57.1%; Score 92; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 7.4e-86;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PNVNNGNLVVDLSTQIFCHNDYPETITDVTYLRGSAVGVLSNFGTVKYSGSSYPP 60
DB 26 PNVNNGNLVVDLSTQIFCHNDYPETITDVTYLRGSAVGVLSNFGTVKYSGSSYPP 85
QY 61 TTSETPRVYVNSRTDKPMPVALYLRPVSSAGG 92
DB 86 TTSETPRVYVNSRTDKPMPVALYLRPVSSAGG 117
RESULT 7
AAE18426
ID AAE18426 standard; Protein; 279 AA.
XX
AC AAE18426;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC60 FliH protein.
XX
KM FliH: Immune response; antibacterial; enterobacillus-related disease;
KM therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli EC60.
XX
FH Key Location/Qualifiers
FH
FT Misc-difference 27 /note= "Encoded by GCC"
FT Misc-difference 66 /note= "Encoded by AGC"
FT Misc-difference 70 /note= "Encoded by AGT"
FT Misc-difference 78 /note= "Encoded by AAT"
FT Misc-difference 119 /note= "Encoded by GCG"
FT Misc-difference 176 /note= "Encoded by CCT"
FT Misc-difference 201 /note= "Encoded by ACC"
FT Misc-difference 269 /note= "Encoded by CAA"
FT Misc-difference 273 /note= "Encoded by GGC"
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMUNE INC.

```

XX  Langermann S, Revel A, Auguste C, Burlein J;
PI  WPI: 2002-171702/22.
XX  N-PSDB; AAD29372.
DR  N-PSDB; AAD29372.
XX  New immunogenic polypeptide, useful as vaccine for protecting against
PT  an enterobacillus-related disease in a patient at risk of contracting
PR  such disease, e.g. urinary tract infection or a bladder infection
PS  Claim 3; Fig 2; 101pp; English.

XX  The invention relates to bacterial immunogenic agents for administration
CC  to humans and non-human animals to stimulate an immune response. The
CC  invention also relates to methods for vaccination of mammalian species
CC  with variants of E. coli FimH protein derived from different strains of
CC  E. coli. The vaccine composition or the antibody is useful for protecting
CC  against and treating an enterobacillus-related disease in a patient
CC  afflicted or at a risk of contracting the disease. In particular, the
CC  disease is a urinary tract or bladder infection. The disease is caused
CC  by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC  The present sequence is Escherichia coli strain EC60 FimH protein.

SQ  Sequence 279 AA;
Query Match 57.1%; Score 92; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 7.4e-86;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PNVNNGONLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLNFGTVKYSGSSYP 60
DB 26 PNVNNGONLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLNFGTVKYSGSSYP 85
OY 61 TTSETPRVVYNSRTDKPWPVALYLPVSSAG 92
DB 86 TTSETPRVVYNSRTDKPWPVALYLPVSSAG 117

RESULT 8
AAE18433
ID AAE18433 standard; Protein; 279 AA.
XX
AC AAE18433;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain J96 FimH protein.
XX
KM FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli J96.
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI: 2002-171702/22.
XX  N-PSDB; AAD29372.
XX  New immunogenic polypeptide, useful as vaccine for protecting against
PT  an enterobacillus-related disease in a patient at risk of contracting
PR  such disease, e.g. urinary tract infection or a bladder infection
PS  Claim 3; Page 90-91; 101pp; English.

```

```

XX  The invention relates to bacterial immunogenic agents for administration
CC  to humans and non-human animals to stimulate an immune response. The
CC  invention also relates to methods for vaccination of mammalian species
CC  with variants of E. coli FimH protein derived from different strains of
CC  E. coli. The vaccine composition or the antibody is useful for protecting
CC  against and treating an enterobacillus-related disease in a patient
CC  afflicted or at a risk of contracting the disease. In particular, the
CC  disease is a urinary tract or bladder infection. The disease is caused
CC  by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC  The present sequence is Escherichia coli strain J96 FimH protein.

SQ  Sequence 279 AA;
Query Match 57.1%; Score 92; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 7.4e-86;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PNVNNGONLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLNFGTVKYSGSSYP 60
DB 26 PNVNNGONLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLNFGTVKYSGSSYP 85
OY 61 TTSETPRVVYNSRTDKPWPVALYLPVSSAG 92
DB 86 TTSETPRVVYNSRTDKPWPVALYLPVSSAG 117

RESULT 9
AAB47073
ID AAB47073 standard; Protein; 296 AA.
XX
AC AAB47073;
XX
DT 08-MAY-2001 (first entry)
XX
DE Immunogenic complex: FimH-linker-FimG N-terminal extension.
XX
KW Donor; pilus protein; pilin; adhesin; vaccine; urinary epithelia;
KW urinary tract infection; enterobacteriaceae.
XX
OS Chimeric - Escherichia coli.
XX
OS Chimeric - Synthetic.
XX
FH Key
FH FT 1..279 Location/Qualifiers
FT /label= "FimH
FT /note= "Pilus protein component"
FT Peptide 280..283
FT /note= "Linker"
FT Peptide 284..296
FT /label= "N-terminal extension of FimG
FT /note= "Donor strand component"
XX
PN WO200104148-A2.
XX
PD 18-JAN-2001.
XX
PF 13-JUL-2000; 2000WO-US19066.
XX
PR 13-JUL-1999; 99US-0143582.
XX  16-JUL-1999; 99US-0144359.
XX  23-FEB-2000; 2000US-0184442.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Hultgren SJ, Pinkner JS, Sauer F, Barnhart M, Waksman G, Knight S;
XX
DR WPI: 2001-138315/14.
XX  Immunogenic complexes and polypeptides for vaccinating against urinary
PT  tract disease, comprises a pilus protein component and a bacterial
PT  chaperone
XX

```


PS Claim 19; Page 81-82; 92pp; English.

XX
CC This sequence represents the immunogenic complex of the invention.
CC The complex comprises a pilus protein component and a donor complement
CC portion as part of the same amino acid sequence attached through an
CC amino acid linker. The linker is composed of a sequence which
CC readily forms a loop such that the donor strand can loop back towards
CC the pilus protein and form an anti-parallel structure. Pilus associated
CC adhesins, such as FimH are relatively conserved proteins among
CC different species and strains of bacteria, therefore vaccines
CC incorporating the FimH antigen exhibit a broad spectrum of
CC protection compared with current pilus-fiber based vaccines. The
CC immunogenic complexes act by disrupting pilus-mediated attachment
CC of E. coli to urinary epithelia and may prevent or retard the
CC development of urinary tract infections. Vaccines containing the
CC complexes are useful for preventing urinary tract disease in a human
CC caused by the bacterium family enterobacteriaceae specifically
CC Escherichia coli and may also be used in treating the disease.
CC
XX
SQ Sequence 296 AA;

Query Match 57.1%; Score 92; DB 22; Length 296;
Best Local Similarity 100.0%; Pred. No. 7.8e-86;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PNVNCGNLVVDLSTQIFCHNDYPERITDVTLRGSAYGVLNFGTVKYGSSYPP 60
DB 26 PNVNCGNLVVDLSTQIFCHNDYPERITDVTLRGSAYGVLNFGTVKYGSSYPP 85
QY 61 TTSETPRVYNSRTDKPWPVALYLTVPSSAG 92
DB 86 TTSETPRVYNSRTDKPWPVALYLTVPSSAG 117

RESULT 10
AAR6745
ID AAR6745 standard; protein; 300 AA.
XX
AC AAR6745;
XX
DT 13-MAR-1996 (first entry)
XX
DE FimH protein derived from E. coli K12 strain PC31.
XX
KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KM FimH; FimG; receptor binding site.
XX
OS Escherichia coli K12 strain PC31.
XX
FH Key
FT Peptide 1..21 Location/Qualifiers
FT Protein 22..300 /note="Signal peptide"
FT /note="Mature FimH"

W09520657-A1.
PN
PD 03-AUG-1995.
XX
PF 27-JAN-1995; 95WO-DK00042.
XX
PR 27-JAN-1994; 94US-0187166.
XX
XX (GXBI-) GX BIOSYSTEMS AS.
PA
PI Hasley DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
XX WPI; 1995-275442/36.
DR
XX Receptor specific bacterial adhesins - useful for targetting active
PT compounds and microbial cells to locations of receptors
PT
PS Example 1; Page 88-89; 152pp; English.

XX
CC This sequence represents the FimH protein from E. coli K12 strain
CC PC31. FimH is located at the tip of the type 1 fimbriae and also
CC intercalated at intervals in the fimbrial organelle. Most forms of
CC the FimH adhesin target, and bind to, oligosaccharide structures
CC containing terminally located alpha-D-mannoside residues. FimH
CC contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison FimA and
CC the minor components FimG and FimG only have 2 cysteine residues.
CC The localisation of the cysteine residues in FimH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FimH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FimH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. This
CC sequence and those given in AAR6763-76 may be used in the production of
CC a variant FimH adhesin which may be useful for targeting active
CC compounds and microbial cells to locations comprising selected receptors
CC to which the adhesins bind.
CC
XX
SQ Sequence 300 AA;

Query Match 57.1%; Score 92; DB 16; Length 300;
Best Local Similarity 100.0%; Pred. No. 7.9e-86;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PNVNCGNLVVDLSTQIFCHNDYPERITDVTLRGSAYGVLNFGTVKYGSSYPP 60
DB 47 PNVNCGNLVVDLSTQIFCHNDYPERITDVTLRGSAYGVLNFGTVKYGSSYPP 106
QY 61 TTSETPRVYNSRTDKPWPVALYLTVPSSAG 92
DB 107 TTSETPRVYNSRTDKPWPVALYLTVPSSAG 138

RESULT 11
AAB16009
ID AAB16009 standard; Protein; 300 AA.
XX
AC AAB16009;
XX
DT 05-OCT-2000 (first entry)
XX
DE E. coli proliferation associated protein sequence SEQ ID NO:367.
XX
KW Escherichia coli; E. coli; proliferation; inhibition; screening;
KM antimicrobial; bacterial growth; antisense therapy; antibacterial.
XX
OS Escherichia coli.
XX
PN W0200044906-A2.
XX
PD 03-AUG-2000.
XX
PF 27-JAN-2000; 2000WO-US02200.
XX
PR 27-JAN-1999; 99US-0117405.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
PI Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr CJ;
XX
XX Yamamoto RT, Xu HH;
DR
XX WPI; 2000-514822/46.
DR
XX N-PSDB; AAA6015.
XX
PT Novel polynucleotides and polypeptides associated with microorganism
PT proliferation, used to identify inhibitors of bacterial growth and
PT proliferation, for use in antisense therapy -
XX
PS Claim 11; Page 274-275; 316pp; English.

CC more than two urogenital infections within one year, has asymptomatic
CC bacteraemia, is a pregnant woman or a diabetic, is immunocompromised, has
CC a human immunodeficiency virus (HIV) infection, has cancer, is in
CC remission from cancer, or is at risk for end stage renal disease. The
CC method is useful for vaccinating a primate against urogenital tract
CC infections, for treating or ameliorating the symptoms of urogenital
CC tract infections, and also for slowing or preventing progression of
CC a urinary tract infection into end stage renal disease. The present
CC sequence represents E. coli FimH protein.

XX Sequence 300 AA;

Query Match 57.1%; Score 92; DB 23; Length 300;
Best Local Similarity 100.0%; Pred. No. 7, 9e-86;

Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PVTNPGNLYVDLSTQIFCHNDYPERITDVTYLTQSGAYGVLNFGTGVKSGSSYPF 60

Db 47 PVTNPGNLYVDLSTQIFCHNDYPERITDVTYLTQSGAYGVLNFGTGVKSGSSYPF 106

Qy 61 TTSETPRVYNSRTDKPWPVALYLTVPVSSAG 92

Db 107 TTSETPRVYNSRTDKPWPVALYLTVPVSSAG 138

RESULT 14

AA847072 standard; Protein; 304 AA.

AC AA847072;

DT 08-MAY-2001 (first entry)

DE Immunogenic complex: FimH-linker-G1 beta-strand of FimC.

KM Donor: pilus protein; pilin; adhesin; vaccine: urinary epithelia;

KW urinary tract infection; enterobacteriaceae.

OS Chimeric - Escherichia coli.

OS Chimeric - Synthetic.

FH Key Location/Qualifiers

FT Protein 1..279 /label= FimH

FT Peptide 280..289 /note= "Pilus protein component"

FT Peptide 290..304 /note= "Linker"

FT /label= G1 beta-strand of FimC

FT /note= "Donor strand component"

PN WO200104148-A2.

PD 18-JAN-2001.

PE 13-JUL-2000; 2000WO-US19066.

PR 13-JUL-1999; 99US-0143582.

PR 16-JUL-1999; 99US-0144359.

PR 23-FEB-2000; 2000US-0184442.

PA (MEDI-) MEDIMMUNE INC.

PI Hultgren SJ, Pinkner JS, Sauer F, Barnhart M, Waksman G, Knight S;

XX WPI; 2001-138315/14.

PT Immunogenic complexes and polypeptides for vaccinating against urinary

PS Chapterone -

PS Claim 19; Page 80-81; 92pp; English.

CC This sequence represents the immunogenic complex of the invention.
CC The complex comprises a pilus protein component and a donor complement
CC portion as part of the same amino acid sequence attached through an
CC amino acid linker. The linker is composed of a sequence which
CC readily forms a loop such that the donor strand can loop back towards
CC the pilus protein and form an anti-parallel structure. Pilus associated
CC adhesins, such as FimH are relatively conserved proteins among
CC different species and strains of bacteria, therefore vaccines
CC incorporating the FimH antigen exhibit a broad spectrum of
CC protection compared with current pilus-fiber based vaccines. The
CC immunogenic complexes act by disrupting pilus-mediated attachment
CC of E. coli to urinary epithelia and may prevent or retard the
CC development of urinary tract infections. Vaccines containing the
CC complexes are useful for preventing urinary tract disease in a human
CC caused by the bacterium family enterobacteriaceae specifically
CC Escherichia coli and may also be used in treating the disease.

XX Sequence 304 AA;

Query Match 57.1%; Score 92; DB 22; Length 304;
Best Local Similarity 100.0%; Pred. No. 8e-86;

Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PVTNPGNLYVDLSTQIFCHNDYPERITDVTYLTQSGAYGVLNFGTGVKSGSSYPF 60

Db 26 PVTNPGNLYVDLSTQIFCHNDYPERITDVTYLTQSGAYGVLNFGTGVKSGSSYPF 85

Qy 61 TTSETPRVYNSRTDKPWPVALYLTVPVSSAG 92

Db 86 TTSETPRVYNSRTDKPWPVALYLTVPVSSAG 117

RESULT 15

AAE18417 standard; Protein; 279 AA.

AC AAE18417;

DT 07-MAY-2002 (first entry)

DE Escherichia coli strain B228 FimH protein.

KM FimH; Immune response; antibacterial; enterobacillus-related disease;

KW therapy; vaccine; urinary tract infection; bladder.

OS Escherichia coli B228.

FH Key Location/Qualifiers

FT Misc-difference 176 /note= "Encoded by CCR"

FT Misc-difference 203 /note= "Encoded by ACC"

PN WO200204496-A2.

PD 17-JAN-2002.

PE 06-JUL-2001; 2001WO-US21525.

PR 07-JUL-2000; 2000US-216750P.

PA (MEDI-) MEDIMMUNE INC.

PI Langermann S, Revel A, Auguste C, Burlin J;

XX WPI; 2002-171702/22.

DR N-PSDB; AAD29356.

PT New immunogenic polypeptide, useful as vaccine for protecting against

PS an enterobacillus-related disease in a patient at risk of contracting

PS Claim 3; Fig 2; 101pp; English.

XX	The invention relates to bacterial immunogenic agents for administration
CC	to humans and non-human animals to stimulate an immune response. The
CC	invention also relates to methods for vaccination of mammalian species
CC	with variants of E. coli FimH protein derived from different strains of
CC	E. coli. The vaccine composition or the antibody is useful for protecting
CC	afflicted or at risk of contracting the disease. In particular, the
CC	disease is a urinary tract or bladder infection. The disease is caused
CC	by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC	The present sequence is Escherichia coli strain B28 FimH protein.
XX	
SQ	Sequence 279 AA;
Query Match	55.9%; Score 90; DB 23; Length 279;
Best Local Similarity	100.0%; Pred. No. 8.3e-84;
Matches 90; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	3 VNWGNLVVDLSQTFCFCHNDYPERITDVTYLORGSAVGVLSNFGTVKVGSSSYPPFTT 62 28 VNWGNLVVDLSQTFCFCHNDYPERITDVTYLORGSAVGVLSNFGTVKVGSSSYPPFTT 87
Qy	63 SETPRVVYNSRTDKPWPVALYTTPVSSAG 92
Db	88 SETPRVVYNSRTDKPWPVALYTTPVSSAG 117
RESULT 16	
ID	AAE18418
AE18418	standard; Protein; 279 AA.
XX	
AC	AAE18418;
DT	07-MAY-2002 (first entry)
XX	
DE	Escherichia coli strain B28 FimH protein.
KW	FimH; immune response; antibacterial; enterobacillus-related disease;
RW	therapy; vaccine; urinary tract infection; bladder.
XX	
OS	Escherichia coli B238.
XX	
Key	Location/Qualifiers
FT	Misc-difference 24 /note= "Encoded by ATT"
FT	Misc-difference 26 /note= "Encoded by CCC"
FT	Misc-difference 176 /note= "Encoded by CCT"
FT	Misc-difference 201 /note= "Encoded by ACC"
FT	Misc-difference 273 /note= "Encoded by GCC"
FT	Misc-difference 274 /note= "Encoded by GTG"
XX	
PX	WO20020496-A2.
XX	
PD	17-JAN-2002.
XX	
PF	06-JUL-2001; 2001WO-US21525.
XX	
PR	07-JUL-2000; 2000US-216750P.
PA	(MEDI-) MEDIMMUNE INC.
XX	
PI	Langermann S, Revel A, Auguste C, Burlein J;
DR	WPI: 2002-171702/22.
N-PsDB:	AAD29357.
XX	
New immunogenic polypeptide,	useful as vaccine for protecting against
An enterobacillus-related disease	in a patient at risk of contracting

PT	such disease, e.g. urinary tract infection or a bladder infection	-
XX		
PS	Claim 3; Fig 2; 101pp; English.	
XX		
CC	The invention relates to bacterial immunogenic agents for administration	
CC	to humans and non-human animals to stimulate an immune response. The	
CC	invention also relates to methods for vaccination of mammalian species	
CC	with variants of E. coli FimH protein derived from different strains of	
CC	E. coli. The vaccine composition or the antibody is useful for protecting	
CC	against and treating an enterobacillus-related disease in a patient	
CC	afflicted or at a risk of contracting the disease. In particular, the	
CC	disease is a urinary tract or bladder infection. The disease is caused	
CC	by a bacterium of the family Enterobacteriaceae, particularly E. coli.	
CC	The present sequence is Escherichia coli strain B238 FimH protein.	
XX		
SO	Sequence	279 AA;
	Query Match	55.9%; Score 90; DB 23; Length 279;
	Best Local Similarity	100.0%; Pred. No. 8.3e-84;
	Matches	90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	3	VNVGNINLVVDLSIOIFCHNDYPETTTDVTYLQRGSAVGGVLSNFGTYKYGSSSYPPPTT 62
DB	28	VNVGNINLVVDLSIOIFCHNDYPETTTDVTYLQRGSAVGGVLSNFGTYKYGSSSYPPPTT 87
OY	63	SETPRVVYNSRTDKPWPVALYLTTPVSSAGG 92
DB	88	SETPRVVYNSRTDKPWPVALYLTTPVSSAGG 117
	RESULT 17	
ID	AAE18422	standard; Protein; 279 AA.
XX	AAE18422;	
AC	AAE18422;	
XX	07-MAY-2002	(first entry)
DT		
XX	Escherichia coli strain EC42 FimH protein.	
DE		
XX	FimH; immune response; antibacterial; enterobacillus-related disease;	
KN	therapy; vaccine; urinary tract infection; bladder.	
KM		
XX	Escherichia coli EC42.	
OS		
XX	WO200204496-A2.	
PN		
XX	17-JAN-2002.	
PD		
XX	06-JUL-2001; 2001WO-US21525.	
PF		
XX	07-JUL-2000; 2000US-216750P.	
PR		
XX	(MEDI-) MEDIMUNE INC.	
PA		
PI	Langermann S; Revel A, Auguste C, Burrel J;	
XX		
DR	WPI: 2002-171702/22.	
XX	N-PSDB: AAD29361.	
PT	New immunogenic polypeptide, useful as vaccine for protecting against	
PI	an enterobacillus-related disease in a patient at risk of contracting	
PT	such disease, e.g. urinary tract infection or a bladder infection	-
PS	Claim 3; Fig 2; 101pp; English.	
XX		
CC	The invention relates to bacterial immunogenic agents for administration	
CC	to humans and non-human animals to stimulate an immune response. The	
CC	invention also relates to methods for vaccination of mammalian species	
CC	with variants of E. coli FimH protein derived from different strains of	
CC	E. coli. The vaccine composition or the antibody is useful for protecting	
CC	against and treating an enterobacillus-related disease in a patient	
CC	afflicted or at a risk of contracting the disease. In particular, the	

CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC42 FimH protein.
XX
SQ Sequence 279 AA:

Query Match 55.9%; Score 90; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 8.3e-84;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VNVGQNLVVDLSTQIFCHNDYPTITDYTLQKSAVGYLSNFGTVKXSGSSYPEPTT 62
DB 28 VNVGQNLVVDLSTQIFCHNDYPTITDYTLQKSAVGYLSNFGTVKXSGSSYPEPTT 87
QY 63 SETPRVYNSRTDKPWPVALYLTTPVSSAGG 92
DB 88 SETPRVYNSRTDKPWPVALYLTTPVSSAGG 117

RESULT 18
AAE18429

ID AAE18429 standard; Protein: 279 AA.

XX AAE18429;

DT 07-MAY-2002 (first entry)

XX Escherichia coli strain EC80 FimH protein.

XX FimH; Immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli EC80.

XX Key Location/Qualifiers

FT Misc-difference 27 /note= "Encoded by GTC"

FT Misc-difference 165 /note= "Encoded by GCT"

FT Misc-difference 171 /note= "Encoded by GCA"

FT Misc-difference 269 /note= "Encoded by ACA"

FT Misc-difference 269 /note= "Encoded by CAA"

PN WO20020496-A2.

PD 17-JAN-2002.

PF 06-JUL-2001; 2001WO-US21525.

PR 07-JUL-2000; 2000US-216750P.

PA (MEDI-) MEDIMMUNE INC.

PI Langermann S, Revel A, Auguste C, Burtlein J;

DR WPI: 2002-171702/22.

DR N-PSDB: AAD29368.

XX New immunogenic polypeptide, useful as vaccine for protecting against
XX an enterobacillus-related disease in a patient at risk of contracting
XX such disease, e.g. urinary tract infection or a bladder infection -
XX
XX Claim 3; Fig 2; 101pp; English.

CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused

CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC80 FimH protein.
XX
SQ Sequence 279 AA:

Query Match 55.9%; Score 90; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 8.3e-84;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VNVGQNLVVDLSTQIFCHNDYPTITDYTLQKSAVGYLSNFGTVKXSGSSYPEPTT 62
DB 28 VNVGQNLVVDLSTQIFCHNDYPTITDYTLQKSAVGYLSNFGTVKXSGSSYPEPTT 87
QY 63 SETPRVYNSRTDKPWPVALYLTTPVSSAGG 92
DB 88 SETPRVYNSRTDKPWPVALYLTTPVSSAGG 117

RESULT 19
AAE18432

ID AAE18432 standard; Protein: 279 AA.

XX AAE18432;

DT 07-MAY-2002 (first entry)

XX Escherichia coli strain G189 FimH protein.

XX FimH; Immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli G189.

XX Key Location/Qualifiers

FT Misc-difference 176 /note= "Encoded by CCA"

FT Misc-difference 201 /note= "Encoded by ACC"

PN WO20020496-A2.

PD 17-JAN-2002.

PF 06-JUL-2001; 2001WO-US21525.

PR 07-JUL-2000; 2000US-216750P.

PA (MEDI-) MEDIMMUNE INC.

PI Langermann S, Revel A, Auguste C, Burtlein J;

DR WPI: 2002-171702/22.

DR N-PSDB: AAD29371.

XX New immunogenic polypeptide, useful as vaccine for protecting against
XX an enterobacillus-related disease in a patient at risk of contracting
XX such disease, e.g. urinary tract infection or a bladder infection -
XX
XX Claim 3; Fig 2; 101pp; English.

CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain G189 FimH protein.
XX
SQ Sequence 279 AA:

```

Query Match          55.9%; Score 90; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 8.3e-84;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VNVGONLVLDLSTQIFCHNDYPETITDVTYLQRGSAVGVLNFSGTVKYSGSSYPPEPT 62
   |||
DB 28 VNVGONLVLDLSTQIFCHNDYPETITDVTYLQRGSAVGVLNFSGTVKYSGSSYPPEPT 87
   |||

QY 63 SETPRVYVNSRTDKPWPVALYLTTPVSSAG 92
   |||
DB 88 SETPRVYVNSRTDKPWPVALYLTTPVSSAG 117
   |||

RESULT 20
ID AAE18436 standard; Protein; 279 AA.
XX
XX AAE18436;
XX
XX 07-MAY-2002 (first entry)
XX
DE Escherichia coli FimH consensus protein.
XX
XX FimH; immune response; antibacterial; enterobacillus-related disease;
XX KM therapy; vaccine; urinary tract infection; bladder.
XX OS Escherichia coli.
XX PN WO200204496-A2.
XX PD 17-JAN-2002.
XX PF 06-JUL-2001; 2001WO-US21525.
XX PR 07-JUL-2000; 2000US-216750P.
XX PA (MEDI-) MEDIMMUNE INC.
XX PI Langemann S, Revel A, Auguste C, Buirlein J;
XX WPI; 2002-171702/22.
XX
XX New immunogenic polypeptide, useful as vaccine for protecting against
XX PT an enterobacillus-related disease in a patient at risk of contracting
XX PT such disease; e.g. urinary tract infection or a bladder infection -
XX
XX Claim 3; Fig 2; 101pp; English.
XX
XX The invention relates to bacterial immunogenic agents for administration
XX CC to humans and non-human animals to stimulate an immune response. The
XX CC invention also relates to methods for vaccination of mammalian species
XX CC with variants of E. coli FimH protein derived from different strains of
XX CC E. coli. The vaccine composition or the antibody is useful for protecting
XX CC against and treating an enterobacillus-related disease in a patient
XX CC afflicted or at a risk of contracting the disease. In particular, the
XX CC disease is a urinary tract or bladder infection. The disease is caused
XX CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
XX CC The present sequence is Escherichia coli FimH consensus protein.
XX
XX Sequence 279 AA;
XX
Query Match          55.9%; Score 90; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 8.3e-84;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VNVGONLVLDLSTQIFCHNDYPETITDVTYLQRGSAVGVLNFSGTVKYSGSSYPPEPT 62
   |||
DB 28 VNVGONLVLDLSTQIFCHNDYPETITDVTYLQRGSAVGVLNFSGTVKYSGSSYPPEPT 87
   |||

QY 63 SETPRVYVNSRTDKPWPVALYLTTPVSSAG 92
   |||
DB 88 SETPRVYVNSRTDKPWPVALYLTTPVSSAG 117
   |||

```

```

RESULT 21
ID AAR76769 standard; protein; 300 AA.
XX
XX AAR76769;
XX
XX 15-MAR-1996 (first entry)
XX
DE FimH protein derived from E. coli clinical isolate CI#3.
XX
XX FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
XX KM FimA; FimF; FimG; receptor binding site.
XX OS Escherichia coli clinical isolate CI#3.
XX
XX Key Location/Qualifiers
XX FT Peptide 1..21
XX FT /note= "Signal peptide"
XX FT Protein 22..300
XX FT /note= "Mature FimH"
XX
XX WO9520657-A1.
XX PN 03-AUG-1995.
XX PD
XX PF 27-JAN-1995; 95WO-DK00042.
XX PR 27-JAN-1994; 94US-0187166.
XX PA (GXBI-) GX BIOSYSTEMS AS.
XX PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX WPI; 1995-275442/36.
XX DR N-PSDB; AAQ93071.
XX
XX Receptor specific bacterial adhesins - useful for targeting active
XX PT compounds and microbial cells to locations of receptors
XX
XX Example 1; Page 44-45; 152pp; English.
XX
XX The sequences given in AAR76763-76 are FimH proteins from various E.
XX CC coli clinical isolates. FimH is located at the tip of the type 1
XX CC fimbriae and also intercalated at intervals in the fimbrial organelle.
XX CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
XX CC structures containing terminally located alpha-D-mannoside residues.
XX CC FimH contains 4 cysteine residues assumed to direct folding of the
XX CC molecule into distinct functional domains. For comparison FimA and
XX CC the minor components FimF and FimG only have 2 cysteine residues.
XX CC The localisation of the cysteine residues in FimH points to a tandem
XX CC arrangement of two ancestral genes. Similar amino acids can be
XX CC found in similar positions in the two halves of the FimH protein. The
XX CC "midway" point is located roughly around residue 150 in the mature
XX CC protein. The two halves or domains of FimH have evolved differently
XX CC with the N-terminal section becoming the domain harbouring the receptor
XX CC binding site, whereas the C-terminal sector became the domain of the
XX CC molecule required for integration into the fimbrial organelle. These
XX CC sequences may be used in the production of a variant FimH adhesin which
XX CC may be useful for targeting active compounds and microbial cells to
XX CC locations comprising selected receptors to which the adhesins bind.
XX
XX Sequence 300 AA;
XX
Query Match          55.9%; Score 90; DB 16; Length 300;
Best Local Similarity 100.0%; Pred. No. 8.9e-84;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VNVGONLVLDLSTQIFCHNDYPETITDVTYLQRGSAVGVLNFSGTVKYSGSSYPPEPT 62
   |||
DB 49 VNVGONLVLDLSTQIFCHNDYPETITDVTYLQRGSAVGVLNFSGTVKYSGSSYPPEPT 108
   |||

QY 63 SETPRVYVNSRTDKPWPVALYLTTPVSSAG 92

```

DB 109 SETPRVYNSRTDKPWPVALYLTPVSSAGG 138

|||||
RESULT 22
AAR76771
ID AAR76771 standard; protein: 300 AA.

AC AAR76771;
XX
XX
DT 15-MAR-1996 (first entry)

DE FlmH protein derived from E. coli clinical isolate KS-54.

KW FlmH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KM FlmH; FlmH; receptor binding site.

OS Escherichia coli clinical isolate KS-54.

XX Key Location/Qualifiers

FT Peptide 1..21

FT Protein /note= "Signal peptide"

FT Protein /note= "Mature FlmH"

XX WO9520657-A1.

XX 03-AUG-1995.

XX 27-JAN-1995; 95WO-DK00042.

XX 27-JAN-1994; 94US-0187166.

XX (GXBI-) GX BIOSYSTEMS AS.

XX Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;

DR WPI: 1995-275442/36.

DR N-PSDB: AAQ93073.

PT Receptor specific bacterial adhesins - useful for targetting active
XX compounds and microbial cells to locations of receptors

XX Example 1; Page 44-45; 152pp: English.

XX The sequences given in AAR76763-76 are FlmH proteins from various E.
CC coli clinical isolates. FlmH is located at the tip of the type 1
CC fimbriae and also intercalated at intervals in the fimbrial organelle.
CC Most forms of the FlmH adhesin target to, and bind to, oligosaccharide
CC structures containing terminally located alpha-D-mannoside residues.
CC FlmH contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison FlmH and
CC the minor components FlmF and FlmG only have 2 cysteine residues.
CC The localisation of the cysteine residues in FlmH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FlmH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FlmH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of a variant FlmH adhesin which
CC may be useful for targetting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.

XX Sequence 300 AA:

Query Match 55.9%; Score 90; DB 16; Length 300;
Best Local Similarity 100.0%; Pred. No. 8.9e-84;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VNVGNLVVDLSQTQICNDYPTITDYVTLQGSAYGVLNFSGTGVKSGSSYPPTT 62
|||||

DB 49 VNVGNLVVDLSQTQICNDYPTITDYVTLQGSAYGVLNFSGTGVKSGSSYPPTT 108
OY 63 SETPRVYNSRTDKPWPVALYLTPVSSAGG 92
|||||
DB 109 SETPRVYNSRTDKPWPVALYLTPVSSAGG 138

RESULT 23
AAR76773
ID AAR76773 standard; protein: 300 AA.

AC AAR76773;
XX
XX
DT 15-MAR-1996 (first entry)

DE FlmH protein derived from E. coli clinical isolate MJ#9-3.

KW FlmH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KM FlmH; FlmH; receptor binding site.

OS Escherichia coli clinical isolate MJ#9-3.

XX Key Location/Qualifiers

FT Peptide 1..21

FT Protein /note= "Signal peptide"

FT Protein /note= "Mature FlmH"

XX WO9520657-A1.

XX 03-AUG-1995.

XX 27-JAN-1995; 95WO-DK00042.

XX 27-JAN-1994; 94US-0187166.

XX (GXBI-) GX BIOSYSTEMS AS.

XX Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;

DR WPI: 1995-275442/36.

DR N-PSDB: AAQ93067.

PT Receptor specific bacterial adhesins - useful for targetting active
XX compounds and microbial cells to locations of receptors

XX Example 1; Page 44-45; 152pp: English.

XX The sequences given in AAR76763-76 are FlmH proteins from various E.
CC coli clinical isolates. FlmH is located at the tip of the type 1
CC fimbriae and also intercalated at intervals in the fimbrial organelle.
CC Most forms of the FlmH adhesin target to, and bind to, oligosaccharide
CC structures containing terminally located alpha-D-mannoside residues.
CC FlmH contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison FlmH and
CC the minor components FlmF and FlmG only have 2 cysteine residues.
CC The localisation of the cysteine residues in FlmH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FlmH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FlmH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of a variant FlmH adhesin which
CC may be useful for targetting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.

XX Sequence 300 AA:

Query Match 55.9%; Score 90; DB 16; Length 300;
Best Local Similarity 100.0%; Pred. No. 8.9e-84;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 3 VNVCNLTVDLSTQIFCHNDYPERITDVTYTLQSGAYGVLNFSGTAKVSGSSYPFPTT 62
   |||||||
Db 49 VNVCNLTVDLSTQIFCHNDYPERITDVTYTLQSGAYGVLNFSGTAKVSGSSYPFPTT 108
   |||||||

QY 63 SETPRVYNSRTRDKPWPVALYLTTPVSSAG 92
   |||||||
Db 109 SETPRVYNSRTRDKPWPVALYLTTPVSSAG 138

RESULT 24
AAR76774
ID AAR76774 standard; protein; 300 AA.
XX
AC AAR76774;
XX
DT 15-MAR-1996 (first entry)
XX
DE FimH protein derived from E. coli clinical isolate MJ#31-3.
XX
KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
   FimA; FimF; FimG; receptor binding site.
XX
OS Escherichia coli clinical isolate MJ#31-3.
XX
FH Key
   Location/Qualifiers
FT Peptide
   1..21
   /note= "Signal peptide"
FT Protein
   22..300
   /note= "Mature FimH"
XX
PN W09520657-A1.
XX
PD 03-AUG-1995.
XX
PE 27-JAN-1995; 95WO-DK0042.
XX
PR 27-JAN-1994; 94US-0187166.
XX
PA (GXBI-) GX BIOSYSTEMS AS.
XX
PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
DR WPI: 1995-275442/36.
XX
DR N-PSDB; AAQ93068.
XX
PT Receptor specific bacterial adhesins - useful for targetting active
   compounds and microbial cells to locations of receptors
XX
PS Example 1; Page 44-45; 152pp; English.
XX
CC The sequences given in AAR76763-76 are FimH proteins from various E.
   coli clinical isolates. FimH is located at the tip of the type 1
   fimbriae and also intercalated at intervals in the fimbrial organelle.
   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
   structures containing terminally located alpha-D-mannoside residues.
   FimH contains 4 cysteine residues assumed to direct folding of the
   molecule into distinct functional domains. For comparison FimA and
   the minor components FimF and FimG only have 2 cysteine residues.
   CC The localisation of the cysteine residues in FimH points to a tandem
   arrangement of two ancestral genes. Similar amino acids can be
   found in similar positions in the two halves of the FimH protein. The
   "midway" point is located roughly around residue 150 in the mature
   protein. The two halves or domains of FimH have evolved differently
   with the N-terminal section becoming the domain harbouring the receptor
   binding site, whereas the C-terminal sector became the domain of the
   molecule required for integration into the fimbrial organelle. These
   CC sequences may be used in the production of a variant FimH adhesin which
   may be useful for targetting active compounds and microbial cells to
   locations comprising selected receptors to which the adhesins bind.
XX
SQ Sequence 300 AA;

```

```

Query Match
Best Local Similarity 100.0%; Score 90; DB 16; Length 300;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VNVCNLTVDLSTQIFCHNDYPERITDVTYTLQSGAYGVLNFSGTAKVSGSSYPFPTT 62
   |||||||
Db 49 VNVCNLTVDLSTQIFCHNDYPERITDVTYTLQSGAYGVLNFSGTAKVSGSSYPFPTT 108
   |||||||

QY 63 SETPRVYNSRTRDKPWPVALYLTTPVSSAG 92
   |||||||
Db 109 SETPRVYNSRTRDKPWPVALYLTTPVSSAG 138

RESULT 25
AAR76763
ID AAR76763 standard; protein; 300 AA.
XX
AC AAR76763;
XX
DT 15-MAR-1996 (first entry)
XX
DE FimH protein derived from E. coli clinical isolate KB21.
XX
KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
   FimA; FimF; FimG; receptor binding site.
XX
OS Escherichia coli clinical isolate KB21.
XX
FH Key
   Location/Qualifiers
FT Peptide
   1..21
   /note= "Signal peptide"
FT Protein
   22..300
   /note= "Mature FimH"
XX
PN W09520657-A1.
XX
PD 03-AUG-1995.
XX
PE 27-JAN-1995; 95WO-DK0042.
XX
PR 27-JAN-1994; 94US-0187166.
XX
PA (GXBI-) GX BIOSYSTEMS AS.
XX
PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
DR WPI: 1995-275442/36.
XX
DR N-PSDB; AAQ93068.
XX
PT Receptor specific bacterial adhesins - useful for targetting active
   compounds and microbial cells to locations of receptors
XX
PS Example 1; Page 44-45; 152pp; English.
XX
CC The sequences given in AAR76763-76 are FimH proteins from various E.
   coli clinical isolates. FimH is located at the tip of the type 1
   fimbriae and also intercalated at intervals in the fimbrial organelle.
   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
   structures containing terminally located alpha-D-mannoside residues.
   FimH contains 4 cysteine residues assumed to direct folding of the
   molecule into distinct functional domains. For comparison FimA and
   the minor components FimF and FimG only have 2 cysteine residues.
   CC The localisation of the cysteine residues in FimH points to a tandem
   arrangement of two ancestral genes. Similar amino acids can be
   found in similar positions in the two halves of the FimH protein. The
   "midway" point is located roughly around residue 150 in the mature
   protein. The two halves or domains of FimH have evolved differently
   with the N-terminal section becoming the domain harbouring the receptor
   binding site, whereas the C-terminal sector became the domain of the
   molecule required for integration into the fimbrial organelle. These
   CC sequences may be used in the production of a variant FimH adhesin which
   may be useful for targetting active compounds and microbial cells to
   locations comprising selected receptors to which the adhesins bind.
XX

```



```

SQ      Sequence      300 AA;
Query Match      55.9%; Score 90; DB 16; Length 300;
Best Local Similarity 100.0%; Pred. No. 8.9e-84;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 VNVGQNLVVDLSQIFCHNDYPETITDYVTLQGSAYGVLSNFGSVKXSGSSYPPT 62
         |||
DB      49 VNVGQNLVVDLSQIFCHNDYPETITDYVTLQGSAYGVLSNFGSVKXSGSSYPPT 108
         |||

QY      63 SETPRVYVNSRTDKPMPVALYLTPVSSAG 92
         |||
DB      109 SETPRVYVNSRTDKPMPVALYLTPVSSAG 138

RESULT 26
AAR76767
ID      AAR76767 standard; protein; 296 AA.
XX
AC      AAR76767;
XX
DT      15-MAR-1996 (first entry)
XX
DE      FlmH protein derived from E. coli clinical isolate CI#10.
XX
KW      FlmH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW      FlmH; FlmG; receptor binding site.
XX
OS      Escherichia coli clinical isolate CI#10.
XX
FH      Key
FH      Peptide      1..21
FT      /note= "Signal peptide"
FT      Protein      22..296
FT      /note= "Mature FlmH"
XX
PN      WO9520657-A1.
XX
PD      03-AUG-1995.
XX
PF      27-JAN-1995; 95WO-DK00042.
XX
PR      27-JAN-1994; 94US-0187166.
XX
PA      (GXBI-) GX BIOSYSTEMS AS.
XX
PI      Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
DR      WPI: 1995-275442/36.
XX
DR      N-PSDB: AAQ93069.
XX
PT      Receptor specific bacterial adhesins - useful for targetting active
PT      compounds and microbial cells to locations of receptors
XX
PS      Example 1; Page 44-45; 152pp; English.
XX
CC      The sequences given in AAR76763-76 are flmH proteins from various E.
CC      coli clinical isolates. FlmH is located at the tip of the type 1
CC      fimbriae and also intercalated at intervals in the fimbrial organelle.
CC      Most forms of the flmH adhesin target to, and bind to, oligosaccharide
CC      structures containing terminally located alpha-D-mannoside residues.
CC      FlmH contains 4 cysteine residues assumed to direct folding of the
CC      molecule into distinct functional domains. For comparison flmH and
CC      the minor components flmF and flmG only have 2 cysteine residues.
CC      The localisation of the cysteine residues in flmH points to a tandem
CC      arrangement of two ancestral genes. Similar amino acids can be
CC      found in similar positions in the two halves of the flmH protein. The
CC      "midway" point is located roughly around residue 150 in the mature
CC      protein. The two halves or domains of flmH have evolved differently
CC      with the N-terminal section becoming the domain harbouring the receptor
CC      binding site, whereas the C-terminal sector became the domain of the
CC      molecule required for integration into the fimbrial organelle. These
CC      sequences may be used in the production of a variant flmH adhesin which

```

```

CC      may be useful for targetting active compounds and microbial cells to
CC      locations comprising selected receptors to which the adhesins bind.
XX
SQ      Sequence      296 AA;
Query Match      55.3%; Score 89; DB 16; Length 296;
Best Local Similarity 100.0%; Pred. No. 9.3e-83;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 VNVGQNLVVDLSQIFCHNDYPETITDYVTLQGSAYGVLSNFGSVKXSGSSYPPT 62
         |||
DB      49 VNVGQNLVVDLSQIFCHNDYPETITDYVTLQGSAYGVLSNFGSVKXSGSSYPPT 108
         |||

QY      63 SETPRVYVNSRTDKPMPVALYLTPVSSAG 91
         |||
DB      109 SETPRVYVNSRTDKPMPVALYLTPVSSAG 137

RESULT 27
AAR76772
ID      AAR76772 standard; protein; 300 AA.
XX
AC      AAR76772;
XX
DT      15-MAR-1996 (first entry)
XX
DE      FlmH protein derived from E. coli clinical isolate U221-3.
XX
KW      FlmH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW      FlmH; FlmG; receptor binding site.
XX
OS      Escherichia coli clinical isolate U221-3.
XX
FH      Key
FH      Peptide      1..21
FT      /note= "Signal peptide"
FT      Protein      22..300
FT      /note= "Mature FlmH"
XX
PN      WO9520657-A1.
XX
PD      03-AUG-1995.
XX
PF      27-JAN-1995; 95WO-DK00042.
XX
PR      27-JAN-1994; 94US-0187166.
XX
PA      (GXBI-) GX BIOSYSTEMS AS.
XX
PI      Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
DR      WPI: 1995-275442/36.
XX
DR      N-PSDB: AAQ93074.
XX
PT      Receptor specific bacterial adhesins - useful for targetting active
PT      compounds and microbial cells to locations of receptors
XX
PS      Example 1; Page 44-45; 152pp; English.
XX
CC      The sequences given in AAR76763-76 are flmH proteins from various E.
CC      coli clinical isolates. FlmH is located at the tip of the type 1
CC      fimbriae and also intercalated at intervals in the fimbrial organelle.
CC      Most forms of the flmH adhesin target to, and bind to, oligosaccharide
CC      structures containing terminally located alpha-D-mannoside residues.
CC      FlmH contains 4 cysteine residues assumed to direct folding of the
CC      molecule into distinct functional domains. For comparison flmH and
CC      the minor components flmF and flmG only have 2 cysteine residues.
CC      The localisation of the cysteine residues in flmH points to a tandem
CC      arrangement of two ancestral genes. Similar amino acids can be
CC      found in similar positions in the two halves of the flmH protein. The
CC      "midway" point is located roughly around residue 150 in the mature
CC      protein. The two halves or domains of flmH have evolved differently
CC      with the N-terminal section becoming the domain harbouring the receptor

```

CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of a variant FimH adhesin which
CC may be useful for targeting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.

SO Sequence 300 AA;

Query Match 49.1%; Score 79; DB 16; Length 300;

Best Local Similarity 100.0%; Pred. No. 1.7e-72; Mismatches 0; Gaps 0;

Matches 79; Conservative 0; Indels 0; Gaps 0;

OY 3 VNVCNVLVVDSTQIFCHNDYPTITDVTYTLQSGAVGVLNFGSVTKYSGSSYPFPTT 62
|||||

Db 49 VNVCNVLVVDSTQIFCHNDYPTITDVTYTLQSGAVGVLNFGSVTKYSGSSYPFPTT 108
|||||

OY 63 SETPRVYVNSRTDKPWPYA 81
|||||

Db 109 SETPRVYVNSRTDKPWPYA 127
|||||

RESULT 28

AAE18420 ID AAE18420 standard; Protein; 280 AA.

XX AAE18420;

DT 07-MAY-2002 (first entry)

XX Escherichia coli strain B242 FimH protein.

XX FimH; immune response; antibacterial; enterobacillus-related disease;

KW therapy; vaccine; urinary tract infection; bladder.

XX Escherichia coli B242.

XX Key Location/Qualifiers

FT Misc-difference 176 /note= "Encoded by CCG"

FT Misc-difference 201 /note= "Encoded by ACC"

FT Misc-difference 279..280 /note= "Encoded by CAA"

XX WO200204496-A2.

PD 17-JAN-2002.

PF 06-JUL-2001; 2001WO-US21525.

PR 07-JUL-2000; 2000US-216750P.

XX (MEDT-) MEDIMUNE INC.

PI Langermann S, Revel A, Auguste C, Burtlein J;

XX WPI: 2002-171702/22.

DR N-PSDB: AAD29359.

XX New immunogenic polypeptide, useful as vaccine for protecting against

PS Claim 3; Fig 2; 101pp; English.

XX The invention relates to bacterial immunogenic agents for administration

CC to humans and non-human animals to stimulate an immune response. The

CC invention also relates to methods for vaccination of mammalian species

CC with variants of E. coli FimH protein derived from different strains of

CC E. coli. The vaccine composition or the antibody is useful for protecting

CC against and treating an enterobacillus-related disease in a patient

CC afflicted or at a risk of contracting the disease. In particular, the

CC disease is a urinary tract or bladder infection. The disease is caused

CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.

CC The present sequence is Escherichia coli strain B242 FimH protein.

XX

SO Sequence 280 AA;

Query Match 46.6%; Score 75; DB 23; Length 280;

Best Local Similarity 100.0%; Pred. No. 2e-68; Mismatches 0; Gaps 0;

Matches 75; Conservative 0; Indels 0; Gaps 0;

OY 18 FCHNDYPTITDVTYTLQSGAVGVLNFGSVTKYSGSSYPFPTTSETPRVYNSRTDKP 77
|||||

Db 43 FCHNDYPTITDVTYTLQSGAVGVLNFGSVTKYSGSSYPFPTTSETPRVYNSRTDKP 102
|||||

OY 78 WPVALLTPVSSAGG 92
|||||

Db 103 WPVALLTPVSSAGG 117
|||||

RESULT 29

AAE18415 ID AAE18415 standard; Protein; 279 AA.

XX AAE18415;

DT 07-MAY-2002 (first entry)

XX Escherichia coli strain B217 FimH protein.

XX FimH; immune response; antibacterial; enterobacillus-related disease;

KW therapy; vaccine; urinary tract infection; bladder.

XX Escherichia coli B217.

XX Key Location/Qualifiers

FT Misc-difference 62 /note= "Encoded by TCG"

FT Misc-difference 70 /note= "Encoded by AAT"

FT Misc-difference 78 /note= "Encoded by AGT"

FT Misc-difference 176 /note= "Encoded by CCG"

FT Misc-difference 201 /note= "Encoded by ACC"

FT Misc-difference 234 /note= "Encoded by ACG"

FT Misc-difference 244 /note= "Encoded by CGG"

XX WO200204496-A2.

PD 17-JAN-2002.

PF 06-JUL-2001; 2001WO-US21525.

PR 07-JUL-2000; 2000US-216750P.

XX (MEDT-) MEDIMUNE INC.

PI Langermann S, Revel A, Auguste C, Burtlein J;

XX WPI: 2002-171702/22.

DR N-PSDB: AAD29354.

XX New immunogenic polypeptide, useful as vaccine for protecting against

PS Claim 3; Fig 2; 101pp; English.

XX The invention relates to bacterial immunogenic agents for administration

CC to humans and non-human animals to stimulate an immune response. The

CC invention also relates to methods for vaccination of mammalian species

CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain B217 FimH protein.

XX
SQ Sequence 279 AA;

Query Match 41.6%; Score 67; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 3,1e-60;

Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 IKAGSLIAVILRLQTNNTNSDDFOFWNNIYANDVVPYTGCDVSARDVVTLPDYRGSV 154
Db 120 IKAGSLIAVILRLQTNNTNSDDFOFWNNIYANDVVPYTGCDVSARDVVTLPDYRGSV 179
|||||

OY 155 PIPPLTVY 161
Db 180 PIPPLTVY 186
|||||

RESULT 30
AAE18416

ID AAE18416 standard; Protein; 279 AA.

XX AAE18416;

AC 07-MAY-2002 (first entry)

DE Escherichia coli strain B223 FimH protein.

KW FimH; Immune response; antibacterial; enterobacillus-related disease;
KM therapy; vaccine; urinary tract infection; bladder.

XX Escherichia coli B223.

XX Key Location/Qualifiers

FT Misc-difference 176 /note= "Encoded by CCR"

FT Misc-difference 201 /note= "Encoded by ACC"

FT WO200204496-A2.

XX 17-JAN-2002.

XX 06-JUL-2001; 2001WO-US21525.

XX 07-JUL-2000; 2000US-216750P.

XX (MEDI-) MEDIMUNE INC.

XX Langermann S, Revel A, Auguste C, Burlein J;

XX WPI; 2002-171702/22.

XX N-PSDB; AAD293355.

XX New immunogenic polypeptide, useful as vaccine for protecting against
XX an enterobacillus-related disease in a patient at risk of contracting
XX such disease, e.g. urinary tract infection or a bladder infection -

XX Claim 3; Fig 2; 101pp; English.

CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused

CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain B223 FimH protein.

XX
SQ Sequence 279 AA;

Query Match 41.6%; Score 67; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 3,1e-60;

Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 IKAGSLIAVILRLQTNNTNSDDFOFWNNIYANDVVPYTGCDVSARDVVTLPDYRGSV 154
Db 120 IKAGSLIAVILRLQTNNTNSDDFOFWNNIYANDVVPYTGCDVSARDVVTLPDYRGSV 179
|||||

OY 155 PIPPLTVY 161
Db 180 PIPPLTVY 186
|||||

RESULT 31
AAE18425

ID AAE18425 standard; Protein; 279 AA.

XX AAE18425;

AC 07-MAY-2002 (first entry)

DE Escherichia coli strain EC58 FimH protein.

KW FimH; Immune response; antibacterial; enterobacillus-related disease;
KM therapy; vaccine; urinary tract infection; bladder.

XX Escherichia coli EC58.

XX Key Location/Qualifiers

FT Misc-difference 176 /note= "Encoded by CCR"

FT Misc-difference 201 /note= "Encoded by ACC"

FT WO200204496-A2.

XX 17-JAN-2002.

XX 06-JUL-2001; 2001WO-US21525.

XX 07-JUL-2000; 2000US-216750P.

XX (MEDI-) MEDIMUNE INC.

XX Langermann S, Revel A, Auguste C, Burlein J;

XX WPI; 2002-171702/22.

XX N-PSDB; AAD29364.

XX New immunogenic polypeptide, useful as vaccine for protecting against
XX an enterobacillus-related disease in a patient at risk of contracting
XX such disease, e.g. urinary tract infection or a bladder infection -

XX Claim 3; Fig 2; 101pp; English.

CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC58 FimH protein.

XX Sequence 279 AA;

Query Match	41.6%;	Score 67;	DB 23;	Length 279;					
Best Local Similarity	100.0%;	Pred. NO. 3.1e-60;							
Matches 67;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
Db	95	IKAGSLIAYLLIRQNNNNSDDFQFVWNIYANDVVPVPTGGCVSARDVTYTLPPYRGSV 154							
	120	IKAGSLIAYLLIRQNNNNSDDFQFVWNIYANDVVPVPTGGCVSARDVTYTLPPYRGSV 179							
OY	155	PIPLTVY 161							
	180	PIPLTVY 186							
Db	155	PIPLTVY 161							
	180	PIPLTVY 186							
RESULT 32									
AAE18427									
ID	AAE18427 standard; Protein; 279 AA.								
XX									
AC	AAE18427;								
XX									
DI	07-MAY-2002 (first entry)								
XX									
DE	Escherichia coli strain EC61 FimH protein.								
XX									
KW	FimH; immune response; antibacterial; enterobacillus-related disease; therapy; vaccine; urinary tract infection; bladder.								
XX									
OS	Escherichia coli EC61.								
XX									
FH	Key Location/Qualifiers								
FT	Misc-difference 3								
FT	/note= "Encoded by TAT"								
FT	Misc-difference 176								
FT	/note= "Encoded by CCT"								
FT	Misc-difference 201								
FT	/note= "Encoded by ACC"								
XX									
PN	WO200204496-A2.								
PD	17-JAN-2002.								
XX									
PF	06-JUL-2001; 2001WO-US21525.								
XX									
PR	07-JUL-2000; 2000US-216750P.								
XX									
PA	(MEDI-) MEDIMUNE INC.								
PI	Langermann S, Revel A, Auguste C, Burlein J;								
XX									
DR	WPI; 2002-171702/22.								
XX									
DR	N-PSDB; AAD29366.								
XX									
PT	New immunogenic polypeptide, useful as vaccine for protecting against an enterobacillus-related disease in a patient at risk of contracting such disease, e.g. urinary tract infection or a bladder infection								
PS	Claim 3; Fig 2; 101pp; English.								
XX									
CC	The invention relates to bacterial immunogenic agents for administration to humans and non-human animals to stimulate an immune response. The invention also relates to methods for vaccination of mammalian species with variants of E. coli FimH protein derived from different strains of E. coli. The vaccine composition or the antibody is useful for protecting against and treating an enterobacillus-related disease in a patient afflicted or at a risk of contracting the disease. In particular, the disease is a urinary tract or bladder infection. The disease is caused by a bacterium of the family Enterobacteriaceae, particularly E. coli. The present sequence is Escherichia coli strain EC61 FimH protein.								
CC									
XX	Sequence 279 AA;								
XX									
Query Match	41.6%;	Score 67;	DB 23;	Length 279;					
Best Local Similarity	100.0%;	Pred. NO. 3.1e-60;							
Matches 67;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					

OY	95	IKRSILAVLLIRCTNNYNDDQEFWNVIYANDVVYPFGGCDVSARDVTYTLPPYRSV	154
Db	120	IKAGSLAVLLIRCTNNYNDDQEFWNVIYANDVVYPFGGCDVSARDVTYTLPPYRSV	179
OY	155	PIPLTVY	161
Db	180	PIPLTVY	186
	RESULT 33		
	AAEI18430		
XX	ID	AAEI18430 standard; Protein; 279 AA.	
XX	AC	AAEI18430;	
XX	DT	07-MAY-2002 (first entry)	
DE		Escherichia coli strain EC89 F1mH protein.	
XX			
KW		F1mH; immune response; antibacterial; enterobacillus-related disease;	
KW		therapy; vaccine; urinary tract infection; bladder.	
XX			
OS		Escherichia coli EC89.	
XX			
FH	Key	Location/Qualifiers	
FH	Misc-difference 176		
FT	/note=	"Encoded by CCN"	
FT	Misc-difference 201		
FT	/note=	"Encoded by ACC"	
FT	Misc-difference 226		
FT	/note=	"Encoded by GCG"	
FT	Misc-difference 227		
FT	/note=	"Encoded by CGC"	
FT	Misc-difference 231		
FT	/note=	"Encoded by GTT"	
FT	Misc-difference 232		
FT	/note=	"Encoded by ATT"	
XX			
PN	WO200204496-A2.		
XX			
PD	17-JAN-2002.		
XX			
PE	06-JUL-2001; 2001MO-US21525.		
XX			
PR	07-JUL-2000; 2000US-216750P.		
XX			
PA	(MEDI-) MEDIMUNE INC.		
PI			
PI	Langermann S, Revel A, Auguste C, Burteln J;		
XX			
DR	WPI: 2002-171702/22.		
XX	N-PSDB; AAD29369.		
XX			
PT	New immunogenic polypeptide, useful as vaccine for protecting against		
PT	an enterobacillus-related disease in a patient at risk of contracting		
PT	such disease, e.g. urinary tract infection or a bladder infection -		
XX			
PS	Claim 3; Fig 2; 101pp; English.		
XX			
CC	The invention relates to bacterial immunogenic agents for administration		
CC	to humans and non-human animals to stimulate an immune response. The		
CC	invention also relates to methods for vaccination of mammalian species		
CC	with variants of E. coli F1mH protein derived from different strains of		
CC	E. coli. The vaccine composition or the antibody is useful for protecting		
CC	against and treating an enterobacillus-related disease in a patient		
CC	afflicted or at a risk of contracting the disease. In particular, the		
CC	disease is a urinary tract or bladder infection. The disease is caused		
CC	by a bacterium of the family Enterobacteriaceae, particularly E. coli.		
CC	The present sequence is Escherichia coli strain EC89 F1mH protein.		
XX			
SO	Sequence 279 AA;		

Db	180	PILPLTVY	186
RESULT 36			
ID	AAEI8435	standard; Protein; 279 AA.	
XX	AAEI8435;		
XX	07-MAY-2002	(first entry)	
XX	Escherichia coli strain G162	FimH protein.	
DE	FimH; immune response; antibacterial; enterobacillus-related disease;		
KM	therapy; vaccine; urinary tract infection; bladder.		
KW	Escherichia coli G162.		
OS			
XX			
PH	Key	Location/Qualifiers	
FT	Misc-difference 176	/note= "Encoded by CPT"	
FT	Misc-difference 201	/note= "Encoded by ACC"	
XX	WO200204496-A2.		
PN	17-JAN-2002.		
XX			
PD	06-JUL-2001; 2001MO-US21525.		
XX	07-JUL-2000; 2000US-216750P.		
PF	(MED1-) MEDIMUNE INC.		
XX	Langermann S, Revel A, Auguste C, Butlein J;		
PI	WPI: 2002-171702/22.		
DR	N-PsDB: AAD29382.		
XX	New immunogenic polypeptide, useful as vaccine for protecting against		
PT	an enterobacillus-related disease in a patient at risk of contracting		
PT	such disease, e.g. urinary tract infection or a bladder infection -		
PS	Claim 3; Page 88-89; 101pp; English.		
XX	The invention relates to bacterial immunogenic agents for administration		
CC	to humans and non-human animals to stimulate an immune response. The		
CC	invention also relates to methods for vaccination of mammalian species		
CC	with variants of E. coli FimH protein derived from different strains of		
CC	E. coli. The vaccine composition or the antibody is useful for protecting		
CC	against and treating an enterobacillus-related disease in a patient		
CC	afflicted or at a risk of contracting the disease. In particular, the		
CC	disease is a urinary tract or bladder infection. The disease is caused		
CC	by a bacterium of the family Enterobacteriaceae, particularly E. coli.		
XX	The present sequence is Escherichia coli strain G162 FimH protein.		
SQ	Sequence	279 AA;	
Query Match	41.6%;	Score 67;	DB 23; Length 279;
Best Local Similarity	100.0%;	Pred. No. 3; Le-60;	
Matches 67; Conservative	0;	Mismatches	0; Indels 0; Gaps 0;
QY	95 IKAGSLIAVLIRQNNYNNSDDFOFWNVIYANDVVPVTGGCDVSARDVTATLPDYGSSV	154	
Dd	120 IKAGSLIAVLIRQNNYNNSDDFOFWNVIYANDVVPVTGGCDVSARDVTATLPDYGSSV	179	
QY	155 PILPLTVY	161	
Dd	180 PILPLTVY	186	
RESULT 37			
AAEI8428			

ID	AAE18428	standard; Protein; 279 AA.
XX		
AC	AAE18428;	
XX		
DT	07-MAY-2002	(first entry)
XX		
DE	Escherichia coli strain EC62 fliH protein.	
XX		
KW	fliH; immune response; antibacterial; enterobacillus-related disease; therapy; vaccine; urinary tract infection; bladder.	
XX		
OS	Escherichia coli EC62.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 176	/note= "Encoded by CCT"
FT	Misc-difference 201	/note= "Encoded by ACC"
XX		
PN	WO20020496-A2.	
PD	17-JAN-2002.	
XX		
PF	06-JUL-2001; 2001WO-US21525.	
XX		
PR	07-JUL-2000; 2000US-216750P.	
XX		
PA	(MEDI-) MEDIMMUNE INC.	
XX		
PI	Langermann S, Revel A, Auguste C, Burlain J;	
XX		
DR	WPI; 2002-171702/22.	
DR	N-PSDB; AAD29367.	
XX		
PT	New immunogenic polypeptide, useful as vaccine for protecting against an enterobacillus-related disease in a patient at risk of contracting such disease, e.g. urinary tract infection or a bladder infection	
XX		
PS	Claim 3; Fig 2; 101pp; English.	
XX		
CC	The invention relates to bacterial immunogenic agents for administration to humans and non-human animals to stimulate an immune response. The invention also relates to methods for vaccination of mammalian species with variants of E. coli fliH protein derived from different strains of E. coli. The vaccine composition or the antibody is useful for protecting CC against and treating an enterobacillus-related disease in a patient CC afflicted or at a risk of contracting the disease. In particular, the CC disease is a urinary tract or bladder infection. The disease is caused by a bacterium of the family Enterobacteriaceae, particularly E. coli. The present sequence is Escherichia coli strain EC62 fliH protein.	
XX		
SQ	Sequence	279 AA;
XX		
Query Match	37.3%;	Score 60; DB 23; Length 279;
Best Local Similarity	100.0%;	Pred. No. 4,7e-53;
Matches	60;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	102	AVLLIRQINNNNSDDFQYWNVIYANNDDVVPYGGCDVSARDVTYLPDYGSGVPIPLTVY 161
DB	127	AVLLIRQTNNTNSDDFQYWNVIYANNDDVVPYGGCDVSARDVTYLPDYGSGVPIPLTVY 186
XX		
RESULT 38		
ID	AAK76768	standard; protein; 300 AA.
XX		
AC	AAK76768;	
XX		
DT	15-MAR-1996	(first entry)
XX		
DE	fliH protein derived from E. coli clinical isolate CSH-50.	
XX		
FIMH	type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;	

```
KM F1mH; F1mF; F1mG; receptor binding site.
XX
OS Escherichia coli clinical isolate CSH-50.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..300
FT /note= "Mature F1mH"
XX
PN MO9520657-A1.
XX
PD 03-AUG-1995.
XX
PF 27-JAN-1995; 95WO-DK00042.
XX
PR 27-JAN-1994; 94US-0187166.
XX
PA (GXBI-) GX BIOSYSTEMS AS.
XX
PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
PI WPI; 1995-275442/36.
XX
PT Receptor specific bacterial adhesins - useful for targeting active
PT compounds and microbial cells to locations of receptors
XX
PS Example 1; Page 44-45; 152pp; English.
XX
CC The sequences given in AAR76763-76 are F1mH proteins from various E.
CC coli clinical isolates. F1mH is located at the tip of the type 1
CC fimbriae and also intercalated at intervals in the fimbrial organelle.
CC Most forms of the F1mH adhesin target to, and bind to, oligosaccharide
CC structures containing terminally located alpha-D-mannoside residues.
CC F1mH contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison F1mA and
CC the minor components F1mF and F1mG only have 2 cysteine residues.
CC The localisation of the cysteine residues in F1mH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the F1mH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of F1mH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of a variant F1mH adhesin which
CC may be useful for targeting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.
XX
SQ Sequence 300 AA;
SQ
Query Match 36.6%; Score 59; DB 16; Length 300;
Best Local Similarity 100.0%; Pred. No. 5.3e-52;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 ORGSAYGVLSNFSCTGVKYSGSSYPPTTSETPRVVYNSRTDKPMPVALYLPVSSAGG 92
DB 80 ORGSAYGVLSNFSCTGVKYSGSSYPPTTSETPRVVYNSRTDKPMPVALYLPVSSAGG 138
RESULT 39
AAR76766
ID AAR76766 standard; protein: 300 AA.
XX
AC AAR76766;
XX
DT 15-MAR-1996 (first entry)
XX
DE F1mH protein derived from E. coli clinical isolate CI#7.
XX
KW F1mH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW F1mH; F1mF; F1mG; receptor binding site.
XX
```

```
OS Escherichia coli clinical isolate CI#7.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..300
FT /note= "Mature F1mH"
XX
PN MO9520657-A1.
XX
PD 03-AUG-1995.
XX
PF 27-JAN-1995; 95WO-DK00042.
XX
PR 27-JAN-1994; 94US-0187166.
XX
PA (GXBI-) GX BIOSYSTEMS AS.
XX
PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
PI WPI; 1995-275442/36.
XX
PT Receptor specific bacterial adhesins - useful for targeting active
PT compounds and microbial cells to locations of receptors
XX
PS Example 1; Page 44-45; 152pp; English.
XX
CC The sequences given in AAR76763-76 are F1mH proteins from various E.
CC coli clinical isolates. F1mH is located at the tip of the type 1
CC fimbriae and also intercalated at intervals in the fimbrial organelle.
CC Most forms of the F1mH adhesin target to, and bind to, oligosaccharide
CC structures containing terminally located alpha-D-mannoside residues.
CC F1mH contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison F1mA and
CC the minor components F1mF and F1mG only have 2 cysteine residues.
CC The localisation of the cysteine residues in F1mH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the F1mH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of F1mH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of a variant F1mH adhesin which
CC may be useful for targeting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.
XX
SQ Sequence 300 AA;
SQ
Query Match 35.4%; Score 57; DB 16; Length 300;
Best Local Similarity 100.0%; Pred. No. 6e-50;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 LVVDLSTQIFCHNDYPERTTDTVTYLRGSAYGVLSNFSCTGVKYSGSSYPPTTSET 65
DB 55 LVVDLSTQIFCHNDYPERTTDTVTYLRGSAYGVLSNFSCTGVKYSGSSYPPTTSET 111
RESULT 40
AAE18423
ID AAE18423 standard; protein: 279 AA.
XX
AC AAE18423;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC45 F1mH protein.
XX
KW F1mH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli EC45.
```

```
XX XX MO200204496-A2.
XX PN
XX PD 17-JAN-2002.
XX PF 06-JUL-2001; 2001MO-US21525.
XX PR 07-JUL-2000; 2000US-216750P.
XX PS (MED1-) MEDIMUNE INC.
XX PI Langermann S, Revel A, Auguste C, Burteln J;
XX DR WPI; 2002-171702/22.
XX DR N-PSDB; AAD29362.
XX PT New immunogenic polypeptide, useful as vaccine for protecting against
XX PT an enterobacillus-related disease in a patient at risk of contracting
XX PT such disease, e.g. urinary tract infection or a bladder infection -
XX PS Claim 3; Fig 2; 101pp: English.
XX PS
XX CC The invention relates to bacterial immunogenic agents for administration
XX CC to humans and non-human animals to stimulate an immune response. The
XX CC invention also relates to methods for vaccination of mammalian species
XX CC with variants of E. coli FimH protein derived from different strains of
XX CC E. coli. The vaccine composition or the antibody is useful for protecting
XX CC against and treating an enterobacillus-related disease in a patient
XX CC afflicted or at a risk of contracting the disease. In particular, the
XX CC disease is a urinary tract or bladder infection. The disease is caused
XX CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
XX CC The present sequence is Escherichia coli strain EC45 FimH protein.
XX SQ Sequence 279 AA;
XX
Query Match 34.8%; Score 56; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 5.9e-49;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 95 IKASLSIAVLILRQTNNYNSDDPQFVWNIYANNVYVPTGSCDVSARDVYVTLDPY 150
Db 120 IKASLSIAVLILRQTNNYNSDDPQFVWNIYANNVYVPTGSCDVSARDVYVTLDPY 175
RESULT 41
AAR76765
ID AAR76765 standard; protein: 300 AA.
XX AC
XX AAR76765;
XX DT 15-MAR-1996 (first entry)
XX DE FimH protein derived from E. coli clinical isolate CI#4.
XX KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
XX KM FimA; FimF; receptor binding site.
XX OS Escherichia coli clinical isolate CI#4.
XX FT Key Location/Qualifiers
XX FT Peptide 1..21
XX FT /note= "Signal peptide"
XX FT Protein 22..300
XX FT /note= "Mature FimH"
XX PN
XX PN MO9520657-A1.
XX PD 03-AUG-1995.
XX PF 27-JAN-1995; 95WO-DK00042.
XX PR 27-JAN-1994; 94US-0187166.
XX PA
```

```
PA (GXBI-) GX BIOSYSTEMS AS.
XX PI
XX PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX DR WPI; 1995-275442/36.
XX DR N-PSDB; AAQ93065.
XX PT Receptor specific bacterial adhesins - useful for targeting active
XX PT compounds and microbial cells to locations of receptors
XX PS Example 1; Page 44-45; 152pp: English.
XX PS
XX CC The sequences given in AAR76763-76 are FimH proteins from various E.
XX CC coli clinical isolates. FimH is located at the tip of the type 1
XX CC fimbriae and also interspersed at intervals in the fimbrial organelle.
XX CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
XX CC structures containing terminally located alpha-D-mannoside residues.
XX CC FimH contains 4 cysteine residues assumed to direct folding of the
XX CC molecule into distinct functional domains. For comparison FimA and
XX CC the minor components FimF and FimG only have 2 cysteine residues.
XX CC The localisation of the cysteine residues in FimH points to a tandem
XX CC arrangement of two ancestral genes. Similar amino acids can be
XX CC found in similar positions in the two halves of the FimH protein. The
XX CC "midway" point is located roughly around residue 150 in the mature
XX CC protein. The two halves or domains of FimH have evolved differently
XX CC with the N-terminal section becoming the domain harbouring the receptor
XX CC binding site, whereas the C-terminal sector became the domain of the
XX CC molecule required for integration into the fimbrial organelle. These
XX CC sequences may be used in the production of a variant FimH adhesin which
XX CC may be useful for targeting active compounds and microbial cells to
XX CC locations comprising selected receptors to which the adhesins bind.
XX SQ Sequence 300 AA;
XX
Query Match 34.8%; Score 56; DB 16; Length 300;
Best Local Similarity 100.0%; Pred. No. 6.3e-49;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 95 IKASLSIAVLILRQTNNYNSDDPQFVWNIYANNVYVPTGSCDVSARDVYVTLDPY 150
Db 141 IKASLSIAVLILRQTNNYNSDDPQFVWNIYANNVYVPTGSCDVSARDVYVTLDPY 196
RESULT 42
AAR76770
ID AAR76770 standard; protein: 300 AA.
XX AC
XX AAR76770;
XX DT 15-MAR-1996 (first entry)
XX DE FimH protein derived from E. coli clinical isolate F-18.
XX KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
XX KM FimA; FimF; FimG; receptor binding site.
XX OS Escherichia coli clinical isolate F-18.
XX FT Key Location/Qualifiers
XX FT Peptide 1..21
XX FT /note= "Signal peptide"
XX FT Protein 22..300
XX FT /note= "Mature FimH"
XX PN
XX PN MO9520657-A1.
XX PD 03-AUG-1995.
XX PF 27-JAN-1995; 95WO-DK00042.
XX PR 27-JAN-1994; 94US-0187166.
XX PA (GXBI-) GX BIOSYSTEMS AS.
```



```

XX PI      Hastly DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX XX
DR WPI, 1995-275442/36.
DR N-PSDB; AAQ93063.
XX
PT Receptor specific bacterial adhesins - useful for targetting active
PT compounds and microbial cells to locations of receptors
XX
PS Example 1; Page 44-45; 152pp; English.
XX
CC The sequences given in AA876763-76 are FimH proteins from various E.
CC coli clinical isolates. FimH is located at the tip of the type 1
CC fimbriae and also intercalated at intervals in the fimbrial organelle.
CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC structures containing terminally located alpha-D-mannoside residues.
CC FimH contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison FimH and
CC the minor components FimF and FimG only have 2 cysteine residues.
CC The localisation of the cysteine residues in FimH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FimH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FimH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of a variant FimH adhesin which
CC may be useful for targetting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.
XX
SQ Sequence 300 AA;
XX
Query Match 34.8%; Score 56; DB 16; Length 300;
Best Local Similarity 100.0%; Pred. No. 6.3e-49;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 95 IKAGSLIAYLILRQTNVNSDDFOFWNIYANNVVPYPTGCGDVARSADVTYTLDPY 150
DB 141 IKAGSLIAYLILRQTNVNSDDFOFWNIYANNVVPYPTGCGDVARSADVTYTLDPY 196
XX
RESULT 43
ABB09458 ID ABB09458 standard; Protein; 408 AA.
XX
AC ABB09458;
XX
DT 01-JUL-2002 (first entry)
XX
DE Fusion protein prty-FimH-prty.
XX
KW Exoproteinase; immunostimulant; vaccine; anchor peptide;
KM FimH; ciliated adhesive factor.
OS Lactobacillus helveticus.
OS Escherichia coli.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT 1..70 /label= signal_peptide
FT /note= "derived from L. helveticus"
FT 71..295 /label= FimH_fragment
FT /note= "derived from E. coli ciliated adhesive factor"
FT 296..408 /label= anchor_peptide
FT /note= "derived from L. helveticus"
XX
JP2002017357-A.
XX
PD 22-JAN-2002.

```

```

XX XX
XX 04-JUL-2000; 2000JP-0202442.
XX
XX 04-JUL-2000; 2000JP-0202442.
XX
XX (CALV) CALPIS SHOKUHN KOGYO KK.
XX
XX WPI: 2002-221706/28.
XX
XX N-PSDB; ABLS2756.
XX
XX New anchor peptide, useful for anchoring protein onto microbe -
XX
XX Example 1; Page 10-11; 12pp; Japanese.
XX
CC The invention relates to an anchor peptide derived from the
CC exoproteinase of Lactobacillus helveticus FERM BP-6060.
CC The activity of compositions of the invention may be described as
CC immunostimulatory. The invention also includes a fusion protein,
CC containing the anchor peptide and a signal peptide fused to a useful
CC protein onto the surface of a microbe. The microbe can be used as a
CC vaccine. The current sequence represents a fusion protein referred to as
CC prty-FimH-prty. This fusion protein contains a fragment of the FimH
CC protein, which is a ciliated adhesive factor from E. coli, arranged
CC between the signal sequence and anchor sequences derived from L.
XX helveticus exoproteinase (see ABB09456 and ABB09457).
XX
SQ Sequence 408 AA;
XX
Query Match 34.8%; Score 56; DB 23; Length 408;
Best Local Similarity 100.0%; Pred. No. 8.5e-49;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 95 IKAGSLIAYLILRQTNVNSDDFOFWNIYANNVVPYPTGCGDVARSADVTYTLDPY 150
DB 187 IKAGSLIAYLILRQTNVNSDDFOFWNIYANNVVPYPTGCGDVARSADVTYTLDPY 242
XX
RESULT 44
AAE18413 ID AAE18413 standard; Protein; 279 AA.
XX
AC AAE18413;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain B210 FimH protein.
XX
KW FimH; Immune response; antibacterial; enterobacillus-related disease;
KM therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli B210.
XX
FH Key Location/Qualifiers
FT 1..141 /note= "Encoded by GGT"
FT MISC-difference 176 /note= "Encoded by CCR"
FT MISC-difference 201 /note= "Encoded by ACC"
FT MISC-difference 203 /note= "Encoded by GGT"
XX
XX WC020204496-A2.
XX
XX 17-JAN-2002.
XX
XX 06-JUL-2001; 2001MO-US21525.
XX
XX 07-JUL-2000; 2000US-216750P.
XX
XX (MEDI-) MEDIMUNE INC.
XX

```

PI Langermann S, Revel A, Auguste C, Burtlein J;
 XX
 DR WPI: 2002-171702/22.
 DR N-PSDB; AAD29352.
 XX
 PT New immunogenic polypeptide, useful as vaccine for protecting against
 PT an enterobacillus-related disease in a patient at risk of contracting
 PT such disease, e.g. urinary tract infection or a bladder infection
 PS
 PS Claim 3; Fig 2; 101pp; English.
 XX
 CC The invention relates to bacterial immunogenic agents for administration
 CC to humans and non-human animals to stimulate an immune response. The
 CC invention also relates to methods for vaccination of mammalian species
 CC with variants of E. coli FimH protein derived from different strains of
 CC E. coli. The vaccine composition or the antibody is useful for protecting
 CC against and treating an enterobacillus-related disease in a patient
 CC afflicted or at a risk of contracting the disease. In particular, the
 CC disease is a urinary tract or bladder infection. The disease is caused
 CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
 CC The present sequence is Escherichia coli strain B210 FimH protein.
 CC
 SQ Sequence 279 AA:
 Query Match 29.8%; Score 48; DB 23; Length 279;
 Best Local Similarity 100.0%; Pred. No. 9.4e-41;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PVTNNGONLVNVLSTQIFCHNDYPTITDYTLQGSAYGGVLSNFSG 48
 Db 26 PVTNNGONLVNVLSTQIFCHNDYPTITDYTLQGSAYGGVLSNFSG 73
 RESULT 45
 AAR76764
 ID AAR76764 standard: protein; 300 AA.
 XX
 AC AAR76764;
 XX
 DT 15-MAR-1996 (first entry)
 XX
 DS FimH protein derived from E. coli clinical isolate C1#12.
 XX
 KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
 KM FimH; FimH; receptor binding site.
 XX
 OS Escherichia coli clinical isolate C1#12.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /note= "Signal peptide"
 FT 22..300
 FT Protein /note= "Mature FimH"
 XX
 PN WO9520657-A1.
 XX
 PD 03-AUG-1995.
 XX
 PF 27-JAN-1995; 95WO-DK00042.
 XX
 PR 27-JAN-1994; 94US-0187166.
 XX
 PA (GXBI-) GX BIOSYSTEMS AS.
 XX
 PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
 DR WPI: 1995-275442/36.
 DR N-PSDB; AAQ93075.
 XX
 PT Receptor specific bacterial adhesins - useful for targetting active
 PT compounds and microbial cells to locations of receptors
 XX
 PS Example 1; Page 44-45; 152pp; English.

XX
 CC The sequences given in AAR76763-76 are FimH proteins from various E.
 CC coli clinical isolates. FimH is located at the tip of the type 1
 CC fimbriae and also intercalated at intervals in the fimbrial organelle.
 CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
 CC structures containing terminally located alpha-D-mannoside residues.
 CC FimH contains 4 cysteine residues assumed to direct folding of the
 CC molecule into distinct functional domains. For comparison FimA and
 CC the minor components FimF and FimG only have 2 cysteine residues.
 CC The localisation of the cysteine residues in FimH points to a tandem
 CC arrangement of two ancestral genes. Similar amino acids can be
 CC found in similar positions in the two halves of the FimH protein. The
 CC "midway" point is located roughly around residue 150 in the mature
 CC protein. The two halves or domains of FimH have evolved differently
 CC with the N-terminal section becoming the domain harbouring the receptor
 CC binding site, whereas the C-terminal sector became the domain of the
 CC molecule required for integration into the fimbrial organelle. These
 CC sequences may be used in the production of a variant FimH adhesin which
 CC may be useful for targetting active compounds and microbial cells to
 CC locations comprising selected receptors to which the adhesins bind.
 XX
 SQ Sequence 300 AA:
 Query Match 28.6%; Score 46; DB 16; Length 300;
 Best Local Similarity 100.0%; Pred. No. 1.1e-38;
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 95 IKAGSLIAVLILRQTNNYNSDDPQFWNTYANNDDVYVPFGGCDVSA 140
 Db 141 IKAGSLIAVLILRQTNNYNSDDPQFWNTYANNDDVYVPFGGCDVSA 166

Search completed: November 28, 2002, 19:03:02
 Job time : 81 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 18:54:51 ; Search time 25 Seconds
(without alignments)
267.108 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186

Perfect score: 161
Sequence: 1 PYYVNGQNLVVDLSTDFCH.....DYTVTLPIDYKSVPIPLTVY 161

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	57.1	300	1 FIMH_ECOLI	P08191 escherichia
2	9	5.6	1	YDEQ_ECOLI	P77588 escherichia
3	7	4.3	120	1 RL34_ARATH	O42351 arabidopsis
4	7	4.3	120	1 RL34_PEA	P40590 pisum sativ
5	7	4.3	120	1 RL34_TORAC	P41098 nicotiana t
6	7	4.3	131	1 MYP2_HUMAN	P02689 homo sapien
7	7	4.3	291	1 YOHQ_BACSU	P54513 bacillus su
8	7	4.3	338	1 YAS4_SCHPO	Q10140 schistosach
9	7	4.3	629	1 CTRL_HUMAN	P30825 homo sapien
10	7	4.3	753	1 YBHI_ECOLI	P75764 escherichia
11	7	4.3	1040	1 YEGN_ECOLI	P76398 escherichia
12	6	3.7	68	1 VPO_BPPRD	P27390 bacterioph
13	6	3.7	85	1 IVB2_BUNMU	P00989 dungaricus mu
14	6	3.7	111	1 YDGC_ECOLI	P52110 escherichia
15	6	3.7	122	1 RPA3_YEAST	P26755 saccharomyc
16	6	3.7	133	1 NBSM_CHICK	P48306 gallus gall
17	6	3.7	137	1 YEDX_ECOS57	O8XB75 escherichia
18	6	3.7	137	1 YEDX_ECOLI	P76341 escherichia
19	6	3.7	142	1 RS9_PYRAE	O8ZY40 pyrobaculum
20	6	3.7	149	1 H2A2_PEA	P40281 pisum sativ
21	6	3.7	150	1 H2A1_PEA	P25470 pisum sativ
22	6	3.7	154	1 NABP_RHOSH	O53177 rhodobacter
23	6	3.7	159	1 LSPA_STRCA	O59835 staphylococ
24	6	3.7	178	1 RL6_BACSU	P46898 bacillus su
25	6	3.7	179	1 PTH_CHLMU	P49607 chlamydia m
26	6	3.7	194	1 RUVA_FUSNM	O8REJ7 fusobacteri
27	6	3.7	209	1 RL3_BUCAI	P57591 buchnera ap
28	6	3.7	213	1 MUC3_HUMAN	O02505 homo sapien
29	6	3.7	215	1 RPB5_KLUMA	O9P4B9 kluyveromyc
30	6	3.7	215	1 RPB5_YEAST	P20434 saccharomyc
31	6	3.7	216	1 ALKB_SALTY	P37462 salmonella
32	6	3.7	217	1 CAT_PROMI	P07641 proteus mir
33	6	3.7	217	1 R10A_SPOPR	O963b6 spodoptera

34	6	3.7	217	1 VT11_YEAST	O04338 saccharomyc
35	6	3.7	218	1 GCHL_HAEIN	P43866 haemophilus
36	6	3.7	219	1 GLNP_ECOLI	P10345 escherichia
37	6	3.7	225	1 VW_P12H	P19647 human parai
38	6	3.7	225	1 WV_P12HT	P23057 human parai
39	6	3.7	231	1 TRYP_PIG	P00761 sus scrofa
40	6	3.7	235	1 SODM_MAIZE	P09233 zea mays (m
41	6	3.7	241	1 C531_ECOLI	P15483 escherichia
42	6	3.7	243	1 SDBF_RAT	P24338 rattus norv
43	6	3.7	250	1 VNST_PPRV	P03516 punta toro
44	6	3.7	251	1 YD37_MYCLE	P53426 mycobacteri
45	6	3.7	256	1 DSBG_PSEAE	O91106 pseudomonas
46	6	3.7	256	1 FRDB_HAEIN	P44893 haemophilus
47	6	3.7	258	1 RT15_MOUSE	O9dc71 mus musculu
48	6	3.7	261	1 FLIR_ECOLI	P33135 escherichia
49	6	3.7	265	1 XTMA_BACSU	P39785 bacillus su
50	6	3.7	273	1 OPSG_ODOVI	O18914 odocoillius
51	6	3.7	273	1 OPSR_CANFA	O18914 canis fami
52	6	3.7	273	1 OPSR_HORSE	O18912 equus cabal
53	6	3.7	277	1 XKDB_BACSU	P39781 bacillus su
54	6	3.7	277	1 Y154_ARCFU	O30083 archaeoglob
55	6	3.7	278	1 YFOL_STRTR	P96051 streptococ
56	6	3.7	298	1 PYRD_LACPL	P7687 lactobacill
57	6	3.7	301	1 NHAR_ECOLI	P10087 escherichia
58	6	3.7	301	1 YMT7_MYCTU	O50687 mycobacteri
59	6	3.7	306	1 STYL_CAEEL	O20024 caenorhabd
60	6	3.7	308	1 SAL_STIAS	O9PYV8 silurus aso
61	6	3.7	309	1 FMT_CLOPE	O8XJ13 clostridium
62	6	3.7	314	1 YDGH_ECOLI	P76177 escherichia
63	6	3.7	315	1 MFT_HUMAN	O9h2d1 homo sapien
64	6	3.7	322	1 MTRH_METKA	O32869 methanopyru
65	6	3.7	330	1 LAID_PYRAB	O9V212 pyrococcus
66	6	3.7	333	1 STLK_DROME	P83098 drosophila
67	6	3.7	333	1 Y4M0_RHITSN	P55569 rhizobium s
68	6	3.7	338	1 G3P_TRIHA	P87197 trichoderma
69	6	3.7	340	1 RL3_HALMA	P20279 halocaula
70	6	3.7	340	1 YDDR_ECOLI	P77308 escherichia
71	6	3.7	341	1 CBRP_RAT	P18576 rattus norv
72	6	3.7	345	1 CAMP_PIG	P00795 sus scrofa
73	6	3.7	346	1 CBRP_MOUSE	P23708 mus musculu
74	6	3.7	346	1 YF20_MYCTU	O50587 mycobacteri
75	6	3.7	347	1 CBRP_HUMAN	P23511 homo sapien
76	6	3.7	350	1 G3P_TREPA	O83816 treponema p
77	6	3.7	350	1 OPSL_CALJA	P34989 callitrich
78	6	3.7	356	1 PROB_AQUAE	O67209 aquifex aeo
79	6	3.7	359	1 PTP4_CABEL	P34442 caenorhabd
80	6	3.7	364	1 OPSG_HUMAN	P04001 homo sapien
81	6	3.7	364	1 OPSR_CAPII	O95170 capra hircu
82	6	3.7	364	1 OPSR_FELCA	O18913 felis silve
83	6	3.7	364	1 OPSR_HUMAN	P04000 homo sapien
84	6	3.7	371	1 THIT_SUISO	O98096 sulfolobus
85	6	3.7	376	1 MID2_YEAST	P36077 saccharomyc
86	6	3.7	391	1 CAR1_CANAL	P28872 candida alb
87	6	3.7	391	1 KC21_CHICK	P21868 gallus gall
88	6	3.7	391	1 KC21_HUMAN	P19138 homo sapien
89	6	3.7	391	1 KC21_MOUSE	O60717 mus musculu
90	6	3.7	391	1 KC21_RABIT	P33634 oryctolagus
91	6	3.7	391	1 KC21_RAT	P19319 rattus norv
92	6	3.7	391	1 YTDJ_BACSU	P80861 bacillus su
93	6	3.7	392	1 KC22_XENLA	P28002 xenopus lae
94	6	3.7	395	1 RRRP_P12H	P23055 human parai
95	6	3.7	395	1 RRRP_P12HT	P23055 human parai
96	6	3.7	397	1 PKG_ZYMOO	P09404 zymomonas m
97	6	3.7	407	1 RL4A_ARATH	P49651 arabidopsis
98	6	3.7	408	1 ODP2_RICPR	O9zd40 riceetisia
99	6	3.7	412	1 SDAC_HAEIN	P44615 haemophilus
100	6	3.7	413	1 GAT1_MOUSE	P17679 mus musculu
101	6	3.7	425	1 MNTB_BACSU	P96593 bacillus su
102	6	3.7	426	1 MOEA_MYCTU	O05577 mycobacteri
103	6	3.7	433	1 ANM2_HUMAN	P55345 homo sapien
104	6	3.7	437	1 ACRO_RAT	P29293 rattus norv
105	6	3.7	437	1 GNTT_ECOLI	P39835 escherichia
106	6	3.7	437	1 INTR_SACER	P22877 saccharopol

253	5	3.1	51	1	RL33_PSEAE	09htng pseudomonas	326	3.1	101	1	VG30_BPM15	005339 mycobacteri
254	5	3.1	54	1	RL33_ECOLI	P02436 escherichia	327	3.1	102	1	CH10_STRAL	Q00769 streptomyce
255	5	3.1	55	1	RL33_VIBCH	09kv27 vibrio chol	328	3.1	102	1	CH10_STRCC	P40172 streptomyce
256	5	3.1	55	1	RL33_YERPE	08zfp1 yersinia pe	329	3.1	102	1	S111_RABIT	P24480 oryctolagus
257	5	3.1	58	1	AKP2_ANRAF	PE1348 antipoleura	330	3.1	102	1	S114_HUMAN	060417 homo sapien
258	5	3.1	61	1	CSRA_PSEAE	069078 pseudomonas	331	3.1	103	1	ES6D_MYCTU	005440 mycobacteri
259	5	3.1	63	1	CE3F_HYPCU	P50722 hyphantria	332	3.1	103	1	HIS2_RHOSH	P50335 rhodobacteri
260	5	3.1	63	1	CE3G_HYPCU	003294 drosophila	333	3.1	103	1	KAC4_RABIT	P01840 oryctolagus
261	5	3.1	63	1	PER_DROI	091544 fowipox vir	334	3.1	103	1	RL24_HAEN	P44362 haemophilus
262	5	3.1	63	1	RPO9_FOMPV	09g8b6 myxoma viru	335	3.1	103	1	VC20_VACCC	P21104 vaccinia v1
263	5	3.1	63	1	RPO9_MYVVL	09g921 Shope fibro	336	3.1	103	1	G1R_YEAST	P53325 saccharomyc
264	5	3.1	63	1	RPO9_SEVKA	089560 vaccinia vi	337	3.1	104	1	PF4V_HUMAN	09uap3 homo sapien
265	5	3.1	63	1	RPO9_VACCV	007047 variola vir	338	3.1	104	1	PLF4_HUMAN	P06765 rattus norv
266	5	3.1	63	1	RPO9_YLBY	09dhg8 yaba-like d	339	3.1	105	1	PLF4_RAT	09zci13 rickettsia
267	5	3.1	63	1	RPO9_YLBY	P72192 pseudomonas	340	3.1	105	1	RL21_RICPR	P31949 homo sapien
268	5	3.1	63	1	TAPB_PSEFR	003295 pseudomonas	341	3.1	105	1	S111_HUMAN	09zci13 rickettsia
269	5	3.1	64	1	CAPA_PSEFR	003295 sulfolobus	342	3.1	105	1	Y756_RICPR	09zci13 rickettsia
270	5	3.1	65	1	PER_DROMO	048028 sulfolobus	343	3.1	105	1	Y744_MYCTU	050689 mycobacteri
271	5	3.1	66	1	RPO9_SULSO	048859 streptomyce	344	3.1	106	1	V547_IBVU4	P30247 avian infec
272	5	3.1	66	1	CSPF_STRCO	P10307 bacterioph	345	3.1	106	1	Y547_RICPR	09zci13 rickettsia
273	5	3.1	67	1	RPOZ_TREPA	003802 bacterioph	346	3.1	107	1	GR1A_RICPR	09zci13 rickettsia
274	5	3.1	67	1	VLVS_BP73	P14108 bacterioph	347	3.1	107	1	ULB1_HCMVA	P16831 human cytom
275	5	3.1	67	1	VLVS_BP77	09z386 rhizobium m	348	3.1	107	1	V5MP_IBVU5	P30248 avian infec
276	5	3.1	67	1	Y7K8_BP722	P31569 synecococc	349	3.1	107	1	YB11_BACHD	08yeg9 bacillus ha
277	5	3.1	69	1	CSPA_RHIME	029416 archaeoglob	350	3.1	107	1	KR11_CHICK	08yeg9 brucella me
278	5	3.1	69	1	PSAE_SYNP2	09m53 escherichia	351	3.1	108	1	RSN_HUMAN	08yeg9 h resistin
279	5	3.1	69	1	Y842_ARCFU	041657 vicinia faba	352	3.1	108	1	YNFA_ECO57	08x7a6 escherichia
280	5	3.1	75	1	MCHB_ECOLI	094559 cicier ariet	353	3.1	108	1	YNFA_SALTI	08x7a6 escherichia
281	5	3.1	77	1	MT2_VICPA	P52368 murine cyto	354	3.1	108	1	YNFA_SALTI	08x7a6 escherichia
282	5	3.1	79	1	MT2_CICAR	046229 clover prol	355	3.1	108	1	YNFA_SALTI	08x7a6 escherichia
283	5	3.1	79	1	PE28_MCMVK	044161 prunus arme	356	3.1	108	1	YNFA_SALTI	08x7a6 escherichia
284	5	3.1	79	1	RL22_CLOP	082221 arbidopsi	357	3.1	108	1	YR44_BACHD	08x7a6 escherichia
285	5	3.1	79	1	RL22_CLOP	P31566 oenothera v	358	3.1	109	1	V5MP_IBVU4	08x7a6 escherichia
286	5	3.1	80	1	RUXG_ARATH	P24715 medicago sa	359	3.1	109	1	V5MP_IBVU4	08x7a6 escherichia
287	5	3.1	80	1	RUXG_ARATH	P38839 arbidopsi	360	3.1	109	1	V5MP_IBVU4	08x7a6 escherichia
288	5	3.1	81	1	RUXG_ARATH	P51427 arbidopsi	361	3.1	110	1	YQ39_MYCTU	09abf9 caulobacter
289	5	3.1	82	1	S61B_MEDSA	P27056 rhodobacter	362	3.1	110	1	RLA2_SCHPO	031553 bacillus su
290	5	3.1	83	1	RS3_ARATH	057952 methanococ	363	3.1	110	1	Y332_METJA	P02400 saccharomyc
291	5	3.1	83	1	YF24_ARCFU	001447 yersinia en	364	3.1	110	1	YQ37_METJA	057778 methanococ
292	5	3.1	85	1	RNFH_RHOCA	P06791 human papil	365	3.1	110	1	YQ39_MYCTU	029041 archaeoglob
293	5	3.1	86	1	Y572_METJA	Q01482 triticum ae	366	3.1	110	1	Y225_METJA	P71938 mycobacteri
294	5	3.1	87	1	Y5CF_YEREN	P31800 alcalligenes	367	3.1	110	1	Y225_METJA	060284 methanococ
295	5	3.1	88	1	WIRA_WHEAT	Q11081 caenorhabd	368	3.1	111	1	LYEG_MOUSE	005717 bos taurus
296	5	3.1	88	1	WIRA_WHEAT	P33722 samia cynth	369	3.1	111	1	LYEG_MOUSE	P18389 rana japoni
297	5	3.1	88	1	WIRA_WHEAT	044893 picea glauc	370	3.1	111	1	LYEG_MOUSE	P33461 mus musculu
298	5	3.1	90	1	YTPC_ALCEU	048879 loofah wite	371	3.1	111	1	YQ39_MYCTU	P22222 acroleptism
299	5	3.1	90	1	YTPC_ALCEU	048879 loofah wite	372	3.1	111	1	YQ39_MYCTU	P53245 saccharomyc
300	5	3.1	91	1	YTPC_ALCEU	048879 loofah wite	373	3.1	111	1	YQ39_MYCTU	068946 azomonas ma
301	5	3.1	91	1	YTPC_ALCEU	048879 loofah wite	374	3.1	111	1	YQ39_MYCTU	P24094 rhodobacteri
302	5	3.1	91	1	YTPC_ALCEU	048879 loofah wite	375	3.1	111	1	YQ39_MYCTU	006285 mycobacteri
303	5	3.1	92	1	YTPC_ALCEU	048879 loofah wite	376	3.1	111	1	YQ39_MYCTU	P3496 pediococcus
304	5	3.1	92	1	YTPC_ALCEU	048879 loofah wite	377	3.1	111	1	YQ39_MYCTU	P48286 thermus the
305	5	3.1	93	1	YTPC_ALCEU	048879 loofah wite	378	3.1	111	1	YQ39_MYCTU	091548 streptomyce
306	5	3.1	94	1	YTPC_ALCEU	048879 loofah wite	379	3.1	111	1	YQ39_MYCTU	091548 streptomyce
307	5	3.1	95	1	YTPC_ALCEU	048879 loofah wite	380	3.1	111	1	YQ39_MYCTU	P75570 mycoplasma
308	5	3.1	95	1	YTPC_ALCEU	048879 loofah wite	381	3.1	111	1	YQ39_MYCTU	P81389 homatius ame
309	5	3.1	95	1	YTPC_ALCEU	048879 loofah wite	382	3.1	111	1	YQ39_MYCTU	046240 clostridium
310	5	3.1	95	1	YTPC_ALCEU	048879 loofah wite	383	3.1	111	1	YQ39_MYCTU	P24020 sus scrofa
311	5	3.1	97	1	YTPC_ALCEU	048879 loofah wite	384	3.1	111	1	YQ39_MYCTU	P34192 crossostoma
312	5	3.1	98	1	YTPC_ALCEU	048879 loofah wite	385	3.1	111	1	YQ39_MYCTU	P44358 haemophilus
313	5	3.1	98	1	YTPC_ALCEU	048879 loofah wite	386	3.1	111	1	YQ39_MYCTU	060262 methanococ
314	5	3.1	99	1	YTPC_ALCEU	048879 loofah wite	387	3.1	111	1	YQ39_MYCTU	060262 methanococ
315	5	3.1	99	1	YTPC_ALCEU	048879 loofah wite	388	3.1	111	1	YQ39_MYCTU	060262 methanococ
316	5	3.1	99	1	YTPC_ALCEU	048879 loofah wite	389	3.1	111	1	YQ39_MYCTU	060262 methanococ
317	5	3.1	99	1	YTPC_ALCEU	048879 loofah wite	390	3.1	111	1	YQ39_MYCTU	060262 methanococ
318	5	3.1	99	1	YTPC_ALCEU	048879 loofah wite	391	3.1	111	1	YQ39_MYCTU	060262 methanococ
319	5	3.1	99	1	YTPC_ALCEU	048879 loofah wite	392	3.1	111	1	YQ39_MYCTU	060262 methanococ
320	5	3.1	100	1	YTPC_ALCEU	048879 loofah wite	393	3.1	111	1	YQ39_MYCTU	060262 methanococ
321	5	3.1	100	1	YTPC_ALCEU	048879 loofah wite	394	3.1	111	1	YQ39_MYCTU	060262 methanococ
322	5	3.1	100	1	YTPC_ALCEU	048879 loofah wite	395	3.1	111	1	YQ39_MYCTU	060262 methanococ
323	5	3.1	101	1	YTPC_ALCEU	048879 loofah wite	396	3.1	111	1	YQ39_MYCTU	060262 methanococ
324	5	3.1	101	1	YTPC_ALCEU	048879 loofah wite	397	3.1	111	1	YQ39_MYCTU	060262 methanococ
325	5	3.1	101	1	YTPC_ALCEU	048879 loofah wite	398	3.1	111	1	YQ39_MYCTU	060262 methanococ

545	3.1	170	1	MIR3_MOUSE	Q9twa3 mus musculus	618	3.1	189	1	APOD_MOUSE	P51910 mus musculus
546	3.1	170	1	OBPA_HUMAN	Q9nys6 homo sapien	619	3.1	189	1	APOD_RAT	P23593 rattus norv
547	3.1	170	1	OBPA_HUMAN	Q9nys6 homo sapien	620	3.1	189	1	COAT_CCMV	P03601 cowpea chlo
548	3.1	171	1	AR20_YEAST	P33204 saccharomyc	621	3.1	189	1	COAT_YEAST	P53600 saccharomyc
549	3.1	171	1	YB74_YEAST	P70790 agrobacteri	622	3.1	190	1	C24A_BISBI	Q95173 b cytochrom
550	3.1	171	1	YB74_YEAST	P70790 agrobacteri	623	3.1	190	1	C24A_BOVIN	O46521 bos taurus
551	3.1	172	1	ILVH_METJA	Q9tly1 cyanidium c	624	3.1	190	1	COAT_BMDV	P15158 belladonna
552	3.1	172	1	ILVH_METJA	Q9tly1 cyanidium c	625	3.1	190	1	SOMA_LAMPA	P37885 lama guanico
553	3.1	172	1	MRED_BACSU	057625 methanococc	626	3.1	190	1	Y516_ABRPE	O9yev5 anoperyum p
554	3.1	172	1	YBBO_ECOLI	001467 bacillus su	627	3.1	190	1	Y516_ABRPE	P25651 saccharomyc
555	3.1	173	1	CD3D_MOUSE	P45470 escherichia	628	3.1	191	1	COX2_RHOSH	P43315 lymentria d
556	3.1	173	1	CHB2_LYMDI	P04235 mus musculus	630	3.1	191	1	COX2_RHOSH	P56340 rhodobacter
557	3.1	173	1	CHB2_LYMDI	P04235 mus musculus	631	3.1	191	1	Y611_METJA	P03182 epstein-bar
558	3.1	173	1	CRGC_HUMAN	P07315 homo sapien	632	3.1	191	1	Y611_METJA	O58028 escherichia
559	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	633	3.1	191	1	Y611_METJA	O46306 escherichia
560	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	634	3.1	192	1	Y611_METJA	O46306 escherichia
561	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	635	3.1	192	1	Y611_METJA	O46306 escherichia
562	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	636	3.1	192	1	Y611_METJA	O46306 escherichia
563	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	637	3.1	192	1	Y611_METJA	O46306 escherichia
564	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	638	3.1	192	1	Y611_METJA	O46306 escherichia
565	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	639	3.1	192	1	Y611_METJA	O46306 escherichia
566	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	640	3.1	192	1	Y611_METJA	O46306 escherichia
567	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	641	3.1	192	1	Y611_METJA	O46306 escherichia
568	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	642	3.1	192	1	Y611_METJA	O46306 escherichia
569	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	643	3.1	192	1	Y611_METJA	O46306 escherichia
570	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	644	3.1	192	1	Y611_METJA	O46306 escherichia
571	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	645	3.1	192	1	Y611_METJA	O46306 escherichia
572	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	646	3.1	192	1	Y611_METJA	O46306 escherichia
573	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	647	3.1	192	1	Y611_METJA	O46306 escherichia
574	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	648	3.1	192	1	Y611_METJA	O46306 escherichia
575	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	649	3.1	192	1	Y611_METJA	O46306 escherichia
576	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	650	3.1	192	1	Y611_METJA	O46306 escherichia
577	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	651	3.1	192	1	Y611_METJA	O46306 escherichia
578	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	652	3.1	192	1	Y611_METJA	O46306 escherichia
579	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	653	3.1	192	1	Y611_METJA	O46306 escherichia
580	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	654	3.1	192	1	Y611_METJA	O46306 escherichia
581	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	655	3.1	192	1	Y611_METJA	O46306 escherichia
582	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	656	3.1	192	1	Y611_METJA	O46306 escherichia
583	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	657	3.1	192	1	Y611_METJA	O46306 escherichia
584	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	658	3.1	192	1	Y611_METJA	O46306 escherichia
585	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	659	3.1	192	1	Y611_METJA	O46306 escherichia
586	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	660	3.1	192	1	Y611_METJA	O46306 escherichia
587	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	661	3.1	192	1	Y611_METJA	O46306 escherichia
588	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	662	3.1	192	1	Y611_METJA	O46306 escherichia
589	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	663	3.1	192	1	Y611_METJA	O46306 escherichia
590	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	664	3.1	192	1	Y611_METJA	O46306 escherichia
591	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	665	3.1	192	1	Y611_METJA	O46306 escherichia
592	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	666	3.1	192	1	Y611_METJA	O46306 escherichia
593	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	667	3.1	192	1	Y611_METJA	O46306 escherichia
594	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	668	3.1	192	1	Y611_METJA	O46306 escherichia
595	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	669	3.1	192	1	Y611_METJA	O46306 escherichia
596	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	670	3.1	192	1	Y611_METJA	O46306 escherichia
597	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	671	3.1	192	1	Y611_METJA	O46306 escherichia
598	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	672	3.1	192	1	Y611_METJA	O46306 escherichia
599	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	673	3.1	192	1	Y611_METJA	O46306 escherichia
600	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	674	3.1	192	1	Y611_METJA	O46306 escherichia
601	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	675	3.1	192	1	Y611_METJA	O46306 escherichia
602	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	676	3.1	192	1	Y611_METJA	O46306 escherichia
603	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	677	3.1	192	1	Y611_METJA	O46306 escherichia
604	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	678	3.1	192	1	Y611_METJA	O46306 escherichia
605	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	679	3.1	192	1	Y611_METJA	O46306 escherichia
606	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	680	3.1	192	1	Y611_METJA	O46306 escherichia
607	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	681	3.1	192	1	Y611_METJA	O46306 escherichia
608	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	682	3.1	192	1	Y611_METJA	O46306 escherichia
609	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	683	3.1	192	1	Y611_METJA	O46306 escherichia
610	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	684	3.1	192	1	Y611_METJA	O46306 escherichia
611	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	685	3.1	192	1	Y611_METJA	O46306 escherichia
612	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	686	3.1	192	1	Y611_METJA	O46306 escherichia
613	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	687	3.1	192	1	Y611_METJA	O46306 escherichia
614	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	688	3.1	192	1	Y611_METJA	O46306 escherichia
615	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	689	3.1	192	1	Y611_METJA	O46306 escherichia
616	3.1	173	1	CRGC_MOUSE	P02559 rattus norv	690	3.1	192	1	Y611_METJA	O46306 escherichia
617	3.1	173	1	CRGC_MOUSE	P02559 rattus norv						

837	5	3.1	236	1	YH76_AQUAE	O67652 aquifex aeo	910	3.1	249	1	Y475_CHLMD	Q9pk10 chlamydia m
838	5	3.1	237	1	DUP1_YEAST	P53174 saccharomyc	911	3.1	249	1	YS91_MYCTU	Q10612 mycobacteri
839	5	3.1	237	1	LECA_DIOGR	P08902 dioclea gra	249	3.1	250	1	FCG3_BOVIN	P79107 bos taurus
840	5	3.1	237	1	PSB6_RAT	P28073 rattus norv	912	3.1	250	1	HCYC_SEPOF	P56624 sepiella offic
841	5	3.1	237	1	PYRE_RALSO	Q8y342 ralsionia s	913	3.1	250	1	Y938_MENTJA	O58348 methanococc
842	5	3.1	237	1	RK1_CYAPA	P48125 cyanophora	914	3.1	250	1	YJ72_CORGL	P40111 corynebacte
843	5	3.1	237	1	RL2_ARCFU	O28357 archaeoglob	915	3.1	250	1	YJ72_CORGL	O92868 listeria in
844	5	3.1	237	1	URDF_STRSL	O5056 streptococc	916	3.1	251	1	HIS6_LISIN	O8y995 listeria mo
845	5	3.1	237	1	YAMM_ECOS7	P58315 escherichia	917	3.1	251	1	HIS6_LISIN	O98477 streptomyces
846	5	3.1	238	1	PSB6_MOUSE	O60692 mus musculu	918	3.1	251	1	HIS6_SMLSO	O33774 sulfolobus
847	5	3.1	238	1	Y035_TREPA	O83078 treponema p	919	3.1	251	1	KIKE_HUMAN	O9p9g3 homo sapien
848	5	3.1	239	1	DHAP_BOVIN	P30907 bos taurus	920	3.1	251	1	TREPA_PYRKO	O9yag9 pyrococcus
849	5	3.1	239	1	PDXU_ANASP	O84080 anabeena sp	921	3.1	251	1	Y116_MYCCE	P47362 mycoplasma
850	5	3.1	239	1	PSB6_HUMAN	P28072 homo sapien	922	3.1	251	1	YAT8_SCHPO	O10153 schizosacch
851	5	3.1	239	1	PSB6_HUMAN	P25971 bacillus su	923	3.1	252	1	HIS6_STPAM	O999w8 staphylococ
852	5	3.1	239	1	RL2_HALMA	P20276 halorcula	924	3.1	252	1	OLP1_LYCES	O41350 lycopersico
853	5	3.1	239	1	SFSA_AGR75	P28971 equine hezp	925	3.1	252	1	TFYE_RHILT	P42727 rhizobium l
854	5	3.1	239	1	UR20_HSVB	P33342 escherichia	926	3.1	252	1	Y455_RICPR	O92d88 chondrus cr
855	5	3.1	239	1	YEHG_ECOLI	O9kpi5 vibrio chol	927	3.1	252	1	ATP6_CHOCR	P05805 bos taurus
856	5	3.1	239	1	YH82_VIBCH	O96178 plasmodi	928	3.1	253	1	CAC3_BOVIN	P10805 escherichia
857	5	3.1	240	1	ISPF_PLAFD	O09675 schizosacch	929	3.1	253	1	ENVY_ECOLI	O87kb8 clostridium
858	5	3.1	240	1	YAO2_SCHPO	O31509 bacillus su	930	3.1	253	1	HIS6_CLOAB	O87885 thermomane
859	5	3.1	240	1	YEI1_BACSU	P17055 rhodobacter	931	3.1	253	1	HIS6_THERN	P33593 escherichia
860	5	3.1	241	1	CRTA_RHOCA	O9x0c7 thermotoga	932	3.1	253	1	RECO_RHITO	O985a3 rhizobium l
861	5	3.1	241	1	HISA_THEMA	P76009 escherichia	933	3.1	253	1	NRKD_ECOLI	P07545 agrobacteri
862	5	3.1	241	1	MITE_ECOLI	P42797 arabidopsis	934	3.1	253	1	VIRG_AGR75	P15913 foxyloxy vir
863	5	3.1	241	1	RT10_ARATH	P21625 spiroplasma	935	3.1	253	1	VBPH_APLCA	P41824 aplysia cal
864	5	3.1	241	1	SPIR_SPIME	O57933 methanococ	936	3.1	253	1	HAZR_HUMAN	P01903 homo sapien
865	5	3.1	241	1	Y513_METJA	P24223 escherichia	937	3.1	254	1	IFZA_THEAC	O91h13 thermoplasm
866	5	3.1	242	1	PDXJ_ECOLI	O82n16 salmonella	938	3.1	254	1	PSA7_RAT	P48004 rattus norv
867	5	3.1	242	1	PDXJ_SALTI	O82n16 salmonella	939	3.1	254	1	HA21_MOUSE	O9esn4 mus musculu
868	5	3.1	242	1	PDXJ_SALTY	P72776 synchocyst	940	3.1	254	1	HA22_MOUSE	P01904 mus musculu
869	5	3.1	242	1	PDXJ_SALTY	P52108 escherichia	941	3.1	255	1	HCD2_DROME	P04224 mus musculu
870	5	3.1	242	1	PDXJ_SALTY	O8zcp4 caenorhabd	942	3.1	255	1	HIS6_NEIMB	O18404 drosophila
871	5	3.1	243	1	RSTA_ECOLI	O8zcp4 yersinia pe	943	3.1	255	1	PLSC_NEICO	O9yvh5 neisseria m
872	5	3.1	243	1	YO27_CAEEL	P34677 homo sapien	944	3.1	255	1	PLSC_NEIMB	O95601 neisseria g
873	5	3.1	244	1	CLDC_HUMAN	P09304 varicella-z	945	3.1	255	1	TRR9_HUMAN	O93u41 neisseria m
874	5	3.1	244	1	LEC_LATSP	P05461 bacterioph	946	3.1	255	1	TRR9_HUMAN	O07011 homo sapien
875	5	3.1	244	1	UL04_VYVD	P16349 lathyrus sp	947	3.1	255	1	TRR9_HUMAN	P313001 escherichia
876	5	3.1	244	1	VSID_BPP4	P056749 homo sapien	948	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
877	5	3.1	244	1	YK91_MYCTU	P09304 varicella-z	949	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
878	5	3.1	245	1	ERM1_ENTFA	P12038 enterococc	950	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
879	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	951	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
880	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	952	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
881	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	953	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
882	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	954	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
883	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	955	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
884	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	956	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
885	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	957	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
886	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	958	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
887	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	959	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
888	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	960	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
889	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	961	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
890	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	962	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
891	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	963	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
892	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	964	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
893	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	965	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
894	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	966	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
895	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	967	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
896	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	968	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
897	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	969	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
898	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	970	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
899	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	971	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
900	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	972	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
901	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	973	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
902	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	974	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
903	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	975	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
904	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	976	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
905	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	977	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
906	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	978	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
907	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	979	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
908	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	980	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
909	5	3.1	245	1	ERM2_ENTFA	P12038 enterococc	981	3.1	255	1	TRR9_HUMAN	O9h144 lucilia cup
							982	3.1	261	1	SMRA_ECOLI	O9h144 lucilia cup

```

983 5 3.1 261 1 SURE_METKA 08ty72 methanopyru
984 5 3.1 261 1 YH86_STNY3 P73335 synchocyst
985 5 3.1 261 1 ZNUB_ECOLI P39832 escherichia
986 5 3.1 262 1 DHBG_CYACA P48933 cyanidulum c
987 5 3.1 262 1 FBOL_BOMO P21828 bombyx mori
988 5 3.1 262 1 ITMB_CHICK O42204 gallus gall
989 5 3.1 262 1 LAT_HUMAN O43561 homo sapien
990 5 3.1 262 1 TRUA_PYROH O58941 pyrococcus
991 5 3.1 262 1 VG38_BP72 P07875 bacterioph
992 5 3.1 262 1 Y096_METJA O57561 methanococ
993 5 3.1 262 1 YAZ3_METJA O38429 methanococ
994 5 3.1 262 1 YLJ4_CADEL P34398 caenorhabd1
995 5 3.1 263 1 MPPL_PHLPR P43213 phleum prat
996 5 3.1 263 1 SPAR_SALTY P40701 salmoneilla
997 5 3.1 264 1 CTD2_RAT O35116 rattus norv
998 5 3.1 264 1 XTLI_PSEBU P49155 pseudomonas
999 5 3.1 264 1 Y902_HAEIN P40701 haemophilus
1000 5 3.1 265 1 HME2_BRARE P09015 brachydanio

```

ALIGNMENTS

```

RESULT 1
FIMH_ECOLI STANDARD: PRT: 300 AA.
ID FIMH_ECOLI
AC 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE FimH protein precursor.
GN FimH OR B4320.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8038337; PubMed=2890981;
RA Klemm P., Christiansen G.;
RT "Three fim genes required for the regulation of length and mediation
of adhesion of Escherichia coli type 1 fimbriae.";
RL Mol. Gen. Genet. 208:439-445(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RC MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=90256291; PubMed=1971261;
RA Krogfelt K.A., Bergmans H., Klemm P.;
RT "Direct evidence that the FimH protein is the mannose-specific
adhesin of Escherichia coli type 1 fimbriae.";
RL Infect. Immun. 58:1995-1998(1990).
CC -1- FUNCTION: INVOLVED IN REGULATION OF LENGTH AND MEDIATION OF
ADHESION OF TYPE 1 FIMBRIAE (BUT NOT NECESSARY FOR THE PRODUCTION
OF FIMBRIAE). ADHESIN RESPONSIBLE FOR THE BINDING TO D-MANNOSE. IT
IS LATERALLY POSITIONED AT INTERVALS IN THE STRUCTURE OF THE TYPE
1 FIMBRIAE. IN ORDER TO INTEGRATE FIMH IN THE FIMBRIAE FIMF AND
FIMG ARE NEEDED.
CC -1- SIMILARITY: THIS PROTEIN EXHIBITS HOMOLOGU WITH THE FIMBRIAL
SUBUNIT PROTEIN FIMB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X05672; CAA29156.1; -
CC DR EMBL: U14003; AAA97216.1; -
CC DR EMBL: AE000502; AAC77276.1; -
CC DR PIR: S09563; S09563
CC DR Ecogen: E610315; fimbH.
CC DR InterPro: IPR000259; Fimbrial.
CC DR Pfam: PF00419; Fimbrial; 1.
CC FimH; Signal: Complete proteome.
CC SIGNAL 1 23
CC CHAIN 24 300 FIMH PROTEIN.
CC CONFLICT 197 197 P -> R (IN REF. 1).
CC CONFLICT 222 222 T -> H (IN REF. 1).
CC FT 300 AA; 31473 MW; 9392045158747D CRC64;
CC SQ
CC Query Match 57.1%; Score 92; DB 1; Length 300;
CC Best Local Similarity 100.0%; Pred. No. 4; le-89;
CC Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 PVNVGQNLVVDISTQIFCHNDYPERITDVTYQRCGSAVGVSNGFTYKSGSSYPP 60
DB 47 PVNVGQNLVVDISTQIFCHNDYPERITDVTYQRCGSAVGVSNGFTYKSGSSYPP 106
QY 61 TTSETPRVYNSRTDKMPVALYLPVSSAG 92
DB 107 TTSETPRVYNSRTDKMPVALYLPVSSAG 138

```

```

RESULT 2
YDEQ_ECOLI STANDARD: PRT: 304 AA.
ID YDEQ_ECOLI
AC P77588;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical fimbrial-like protein ydeQ precursor.
GN YDEQ OR B1502.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RC MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasubraman S., Tagami H., Takada J.,
RA Takekoshi K., Takeuchi Y., Wada C., Yamamoto Y., Horinouchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -1- SIMILARITY: TO TYPE-1 FIMBRIAL SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: AE000247; AAC74575.1; -
DR EMBL: D90792; BAA15175.1; -
DR EMBL: D90793; BAA15183.1; -
DR Ecocore: EG13799; ydeq.
DR InterPro: IPR00259; Fimbril.
DR Pfam: PF00419; Fimbril; 1.
KW Hypothetical protein; Fimbril; Signal; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 304
SQ SEQUENCE 304 AA: 32069 MW: 8153C86E3087D99A CRC64;
HYPOTHETICAL FIMBRIL-LIKE PROTEIN YDEQ.

Query Match 5.6%; Score 9; DB 1; Length 304;
Best Local Similarity 100.0%; Pred. No. 0.096;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 GONLVVDLS 14
DB 53 GONLVVDLS 61

RESULT 3
RL34_ARATH STANDARD; PRT; 120 AA.
ID RL34_ARATH
AC 042351;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60S ribosomal protein L34.
GN RPL34 OR ATG26880 OR T2P11.7.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetska I., Kurtz D.B., Kvan A., Lam B.,
RA Lang-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malt R., Marzalli A.,
RA Millscher J., Miranda M., Nguyen M., Nleman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salberg S.L., Schwartz J.R., Shin P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uteback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana";
RL Nature 408:816-820(2000).
CC -1- SIMILARITY: BELONGS TO THE L34E FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL: F20073; CAA23390.1; -
DR EMBL: AC005508; AAD14494.1; -
DR InterPro: IPR001284; Ribosomal_L34E.
DR Pfam: PF01199; Ribosomal_L34E; 1.
DR PRINTS: PR01250; RIBOSOMAL_L34E.
DR PRODOM: PD005148; RIBOSOMAL_L34E; 1.
DR PROSITE: PS01145; RIBOSOMAL_L34E; 1.
KW Ribosomal protein.
SQ SEQUENCE 120 AA: 13705 MW: 5B2334C5C22A44C3 CRC64;
5B2334C5C22A44C3 CRC64;

Query Match 4.3%; Score 7; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 AYGGVLS 44
DB 77 AYGGVLS 83

RESULT 4
RL34_PEA STANDARD; PRT; 120 AA.
ID RL34_PEA
AC P40590;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 60S ribosomal protein L34.
GN RPL34.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eurosid I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxId=3889;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alaska;
RA MEDLINE=96046745; PubMed=7579177;
RA Devitt M.L., Staftrom J.P.;
RT "Cell cycle regulation during growth-dormancy cycles in pea axillary
buds";
RL Plant Mol. Biol. 29:255-265(1995).
CC -1- SIMILARITY: BELONGS TO THE L34E FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: U10047; AAA86953.1; -
DR InterPro: IPR001284; Ribosomal_L34E.
DR Pfam: PF01199; Ribosomal_L34E; 1.
DR PRINTS: PR01250; RIBOSOMAL_L34E.
DR PRODOM: PD005148; RIBOSOMAL_L34E; 1.
DR PROSITE: PS01145; RIBOSOMAL_L34E; 1.
KW Ribosomal protein.
SQ SEQUENCE 120 AA: 13822 MW: 822C8A93589E15ED CRC64;
822C8A93589E15ED CRC64;

Query Match 4.3%; Score 7; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 AYGGVLS 44
DB 77 AYGGVLS 83

RESULT 5
RL34_TOBAC

```

ID RL34_TOBAC STANDARD; PRT: 120 AA.
AC P41098;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE 60S ribosomal protein L34.
GN RPL34.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94355650; PubMed=8075394;
RA Gao J., Kim S.R., Chung Y.Y., Lee J.M., An G.;
RT "Developmental and environmental regulation of two ribosomal protein
RT genes in tobacco.";
RL Plant Mol. Biol. 25:761-770(1994).
CC -I- SIMILARITY: BELONGS TO THE L34E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L27107; AAA57159.1; -
DR EMBL: L27089; AAA57158.1; -
DR InterPro: IPR001284; Ribosomal_L34E.
DR Pfam: PF01199; Ribosomal_L34E; 1.
DR PRINTS: PR01250; Ribosomal_L34.
DR PRODOM: PD005148; Ribosomal_L34E; 1.
DR PROSITE: PS01145; RIBOSOMAL_L34E; 1.
KM Ribosomal protein.
SQ SEQUENCE 120 AA; 13752 MW; 5A987E20093EF6E9 CRC64;

Query Match 4.3%; Score 7; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 AYGVLVS 44
|111111|
DB 77 AYGVLVS 83

RESULT 6
MYP2_HUMAN STANDARD; PRT: 131 AA.
ID MYP2_HUMAN
AC P02689;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myelin P2 protein.
GN PMP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92068191; PubMed=1720307;
RA Hayasaka K., Nanao K., Tahara M., Sato W., Takada G., Miura M.,
RA Uyemura K.;
RT "Isolation and sequence determination of cDNA encoding P2 protein of
RT human peripheral myelin.";
GN Biochem. Biophys. Res. Commun. 181:204-207(1991).
RL [2]
RN RP SEQUENCE.
RA MEDLINE=83058785; PubMed=6183401;
RA Suzuki M., Kitamura K., Sakamoto Y., Uyemura K.;

```

```

RT RT "The complete amino acid sequence of human P2 protein.";
RL J. Neurochem. 39:1759-1762(1982).
RN [3]
RP SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=95054012; PubMed=7525873;
RA Narayanan V., Ripepi B., Jabs E.W., Hawkins A., Griffin C.,
RA Tennekoon G.;
RT "Partial structure and mapping of the human myelin P2 protein gene.";
RL J. Neurochem. 63:2010-2013(1994)
CC -I- FUNCTION: THIS PROTEIN MAY BE A LIPID TRANSPORT PROTEIN IN SCHWANN
CC CELLS.
CC -I- MISCELLANEOUS: P2 PROTEIN AND MYELIN BASIC PROTEIN TOGETHER
CC CONSTITUTE A MAJOR FRACTION OF PERIPHERAL NERVOUS SYSTEM MYELIN
CC PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE FABP/P2/CRBP/CRAB FAMILY OF
CC TRANSPORTERS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D16181; BAA03726.1; -
DR EMBL: D16179; BAA03726.1; JOINED.
DR EMBL: D16180; BAA03726.1; JOINED.
DR EMBL: X62167; CAA44096.1; -
DR EMBL: S73470; AAB32592.2; -
DR EMBL: S73468; AAB32592.2; JOINED.
DR EMBL: S73469; AAB32592.2; JOINED.
DR PIR: A03143; MPM02.
DR PIR: JTO977; JTO977.
DR HSSP: P02690; PMP2.
DR GeneW: HGNC:9117; PMP2.
DR MIM: 170715; -
DR InterPro: IPR000463; Fatty_acid_BP.
DR InterPro: IPR000566; Lipocln_cytfabp.
DR Pfam: PF00061; lipocalinBP.
DR PRINTS: PR00178; FATTYACIDBP.
DR PROSITE: PS00214; FABP; 1.
DR PROSITE: PS00214; FABP; 1.
KM Myelin; lipid-binding; Transport; Acetylation.
FT INIT_MET 0
FT MOD_RES 1 1 ACETYLATION.
FT DISULFD 117 124
FT CONFLICT 24 24 G -> GG (IN REF. 3).
FT CONFLICT 98 98 D -> N (IN REF. 2).
FT CONFLICT 110 110 N -> D (IN REF. 2).
SQ SEQUENCE 131 AA; 14778 MW; 3D0E640BA03E69D2 CRC64;

Query Match 4.3%; Score 7; DB 1; Length 131;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 VTLGRGS 37
|111111|
DB 84 VTLGRGS 90

RESULT 7
YOH0_BACSU STANDARD; PRT: 291 AA.
ID YOH0_BACSU
AC P54513;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yphO.
GN YOH0.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]

```

RP SEQUENCE FROM N.A.
 RC STRAIN-168 / JH642;
 RX MEDLINE-97124195; PubMed-8969508;
 RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
 RA Kobayashi Y.,
 RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
 the *Bacillus subtilis* genome containing the skin element and many
 sporulation genes."
 RL Microbiology 142:3103-3111(1996).
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE-98044033; PubMed-9384377;
 RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
 RA Acedo V., Bertero M.G., Bessieres P., Bolotin A., Borcherdt S.,
 RA Bories R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel N.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Enllich S.D., Emerson P.T.,
 RA Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Filtz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Heaut A.,
 RA Hilbert H., Holtsappel S., Hosono S., Huilo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Kleier-Blanchard M., Klein C.,
 RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Meisel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogawara A., Oudaga B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Potwolk S., Prescott A.M.,
 RA Pesean E., Puig P., Punelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadleir Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter E., Scofield F.,
 RA Sekiguchi J., Sekoyka A., Seror S.J., Serron P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,
 RA Terasaki V., Uchiyama S., Vandenberg M., Vannier F., Vassart A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Welternegger T.,
 RA Winters P., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*."
 RL Nature 390:249-256(1997).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D84433; BAA12552.1; -
 DR EMBL; 299116; CAB14382.1; -
 DR Subtilist; BG11703; yghO.
 DR InterPro; IPR002641; Patatin.
 DR Pfam; PF01734; Patatin; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 34 50 POTENTIAL.
 SO SEQUENCE 291 AA; 32860 MW; B0A7C5DDE4AE37B8 CRC64;

Query Match 4.3%; Score 7; DB 1; Length 291;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 40 GGVLSNF 46
 |||||
 Db 184 GGVLSNF 190

RESULT 8
 YAS4_SCHPO

ID YAS4_SCHPO STANDARD; PRT; 338 AA.
 AC 010140;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C3H8.04 in chromosome I.
 GN SPAC3H8.04.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RX MEDLINE-21848401; PubMed-11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkhardt G., Aert R., Robben J., Gymnopoulos B.,
 RA Wellens J., Ventrone E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritze C., Holzer C., Holzer C., Hillbert H.,
 RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cailen E., Dreano S., Gloux S., Lelure V., Kottler S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas R., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Dague R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Carrutell L., Lowe T., McCombie M.R., Paulsen O., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.,
 RT "The genome sequence of *Schizosaccharomyces pombe*."
 RL Nature 415:871-880(2002).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; 269086; CAA33161.1; -
 DR Hypothetical protein
 KW SEQUENCE 338 AA; 38474 MW; 5B840B5117F9E774 CRC64;

Query Match 4.3%; Score 7; DB 1; Length 338;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 104 LILROTN 110
 |||||
 Db 237 LILROTN 243

RESULT 9
 CTRL_HUMAN
 ID CTRL_HUMAN STANDARD; PRT; 629 AA.
 AC P30825;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE High-affinity cationic amino acid transporter-1 (CAT1) (System
 Y+ basic amino acid transporter) (Ecotropic retroviral leukemia

DE receptor homolog (ERR) (Ecotropic retrovirus receptor homolog).
 GN SLC7A1 OR ATRC1 OR RECL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NC NCB1_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9204065; PubMed=1718082;
 RA Yoshimoto T., Yoshimoto E., Meruelo D.;
 RT "Molecular cloning and characterization of a novel human gene
 homologous to the murine ecotropic retroviral receptor";
 RL Virology 185:10-17(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Urinary bladder;
 RX MEDLINE=92217962; PubMed=1348489;
 RA Albritton L.M., Bowcock A.M., Eddy R.L., Morton G.C., Tseng L.,
 RT Farrer L.A., Cavalli-Sforza L.L., Shows T.B., Cunningham J.M.;
 RT "The human cationic amino acid transporter (ATRC1): physical and
 genetic mapping to 13q12-q14.";
 RL Genomics 12:430-434(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Kamath S.G., Furesz T.C., May B.A., Smith C.H.;
 RT "Cloning, expression, and characterization of a cationic amino acid
 transporter from cultured human placental trophoblast cells.";
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: HIGH-AFFINITY, LOW CAPACITY PERMEASE INVOLVED IN THE
 TRANSPORT OF THE CATIONIC AMINO ACIDS (ARGININE, LYSINE AND
 ORNITHINE) IN NON-HEPATIC TISSUES. MAY ALSO FUNCTION AS AN
 ECOTOPIC RETROVIRAL LEUKEMIA RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: URINARY TISSUES.
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MANY AMINO ACID
 PERMEASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X59155; CAA41869.1; -;
 DR EMBL: X57303; CAA40560.1; -;
 DR EMBL: AF078107; AAC27721.1; -;
 DR Gene: HGNC:11057; SLC7A1;
 DR MIM: 104615; -;
 DR InterPro: IPR002293; AA/rel-permease.
 DR InterPro: IPR004755; Cat_AA-permease.
 DR InterPro: IPR004841; Permease.
 DR Pfam: PF00324; aa-permeases; 1.
 DR TIGRfam: TIGR00906; 2A0303; 1.
 KW Transmembrane; Glycoprotein; Transport; Amino-acid transport;
 KW Receptor.
 FT DOMAIN 1 35 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 36 57 POTENTIAL.
 FT DOMAIN 58 61 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 62 82 POTENTIAL.
 FT DOMAIN 83 102 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 103 123 POTENTIAL.
 FT DOMAIN 124 162 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 163 183 POTENTIAL.
 FT DOMAIN 184 191 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 192 212 POTENTIAL.
 FT DOMAIN 213 246 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 247 267 POTENTIAL.
 FT DOMAIN 268 287 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 288 307 POTENTIAL.
 FT DOMAIN 308 337 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 338 358 POTENTIAL.
 FT DOMAIN 359 384 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 385 405 POTENTIAL.
 FT DOMAIN 406 408 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 409 429 POTENTIAL.
 FT DOMAIN 430 492 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 493 513 POTENTIAL.
 FT DOMAIN 514 526 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 527 551 POTENTIAL.
 FT DOMAIN 552 559 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 560 580 POTENTIAL.
 FT DOMAIN 581 584 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 585 605 POTENTIAL.
 FT DOMAIN 606 629 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 235 235 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 23 23 R -> P (IN REF. 2). CRC64;
 SQ SEQUENCE 629 AA; 67638 MW; 717734d4793647c5 CRC64;
 Query Match 4.3%; Score 7; DB 1; Length 629;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 99 SLIAYLI 105
 Db 501 SLIAYLI 507
 RESULT 10
 YBHJ_ECOLI STANDARD; PRT; 753 AA.
 ID YBHJ_ECOLI
 AC P75764; Q9R755;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein ybhJ.
 GN YBHJ OR B0771.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC NCB1_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IFM ISOMERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

```

CC -----
CC EMBL: AE000179; AAC73858.1; ALT_INIT.
CC EMBL: D90715; BAA35435.1; ALT_INIT.
CC Ecogene: EG13666; ybhJ.
CC InterPro: IPR001030; Aconitase_N.
CC Pfam: PF00330; Aconitase_1.
CC PROSITE: PS00450; ACONITASE_N; 1.
CC PROSITE: PS01244; ACONITASE_2; FALSE_NEG.
CC Hypothetical protein: Iron-sulfur; Complete proteome.
CC METAL 360 360 IRON (IRON-SULFUR CLUSTER) (BY
CC METAL 421 421 SIMILARITY).
CC METAL 421 421 IRON (IRON-SULFUR CLUSTER) (BY
CC METAL 424 424 SIMILARITY).
CC METAL 424 424 IRON (IRON-SULFUR CLUSTER) (BY
CC SEQUENCE 753 AA; 81504 MW; A75DLAE0F7C0F588 CRC64;
CC -----
Query Match
Best Local Similarity 4.3%; Score 7; DB 1; Length 753;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 95 IKAGSLI 101
DB 739 IKAGSLI 745
-----
RESULT 11
YEGN_ECOLI
ID YEGN_ECOLI STANDARD; PRT: 1040 AA.
AC P76398; 008005;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yegn.
GN YEGN OR B2075.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blatter F.R., Plunkett G., IIT, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Mocomma K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horinouchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -1- FUNCTION: COULD BE A DRUG EFFLUX PUMP (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE ACRB/ACRD/ACRF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC -----
CC entries requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000297; AAC75136.1;
CC EMBL: D90845; BAA15929.1;
CC EMBL: D90846; BAA15931.1;
CC Ecogene: EG14057; yegN.
CC InterPro: IPR001036; AcrIflv1n_res.
CC Pfam: PF00873; ACR_tran; 1.
CC PRINTS: PR00702; ACRIFLAVINRP.
CC Hypothetical protein: Transmembrane; Inner membrane; Transport;
CC Complete proteome.
CC DOMAIN 1 20
CC TRANSSEM 21 39
CC TRANSSEM 40 342
CC TRANSSEM 343 362
CC TRANSSEM 363 368
CC TRANSSEM 369 388
CC TRANSSEM 389 394
CC TRANSSEM 395 416
CC TRANSSEM 417 443
CC TRANSSEM 444 462
CC TRANSSEM 463 475
CC TRANSSEM 476 498
CC TRANSSEM 499 536
CC TRANSSEM 537 555
CC TRANSSEM 556 865
CC TRANSSEM 866 885
CC TRANSSEM 886 891
CC TRANSSEM 892 911
CC TRANSSEM 912 917
CC TRANSSEM 918 939
CC TRANSSEM 940 967
CC TRANSSEM 968 986
CC TRANSSEM 987 999
CC TRANSSEM 1000 1022
CC TRANSSEM 1023 1040
CC SEQUENCE 1040 AA; 112077 MW; 195ED989C7AC6E0 CRC64;
CC -----
Query Match
Best Local Similarity 4.3%; Score 7; DB 1; Length 1040;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 99 SLIAYLI 105
DB 450 SLIAYLI 456
-----
RESULT 12
VPO_BPPRD
ID VPO_BPPRD STANDARD; PRT: 68 AA.
AC P27390;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Protein O (GPO).
GN O.
OS Bacteriophage PRD1.
OC Viruses; dsDNA viruses, no RNA stage; Tectiviridae; Tectivirus.
OC NCBI_TaxID=10658;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 61-68.
RX MEDLINE=91306449; PubMed=1853567;
RA Bamford J.K.H., Haenninen A.-L., Pakula T.M., Ojala P.M.,
RA Kalkkinnen N., Frilander M., Bamford D.H.;
RT "Genome organization of membrane-containing bacteriophage PRD1.";
RL Virology 183:658-676(1991).
CC -1- SUBCELLULAR LOCATION: Membrane-associated (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M69077; AAA32470.1; -

DR PIR: G36776; WMBPUB.

SO SEQUENCE 68 AA; 6695 MW; 866351A8AF13746A CRC64;

Query Match 3.7%; Score 6; DB 1; Length 68;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 VIKAGS 99
|||||
Db 34 VIKAGS 39

RESULT 13
ID IVB2_BUNMU STANDARD: PRT; 85 AA.

AC P00989; 042299;
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 40, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-2 bungarotoxin B chain precursor (Bungarotoxin, B2 chain).
OS Bungarus multicinctus (Many-banded krait).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Bungarinae; Bungarus.
ON NCBITaxid=8616;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA MEDLINE=98359733; PubMed=9693106;
RX Wu P.F., Wu S.N., Chang C.C., Chang L.S.;
RT "Cloning and functional expression of B chains of beta-bungarotoxins
from Bungarus multicinctus (Taiwan banded krait).";
RL Biochem. J. 334:87-92(1998).
RN [2]
RP SEQUENCE OF 25-85.
RC TISSUE=Venom;
RA MEDLINE=82239269; PubMed=7096304;
RX Kondo K., Toda H., Marita K., Lee C.Y.;
RT "Amino acid sequence of beta 2 bungarotoxin from Bungarus
multicinctus venom. The amino acid substitutions in the B chains.";
RL J. Biochem. 91:1519-1530(1982).
CC -I- FUNCTION: BETA-2 BUNGAROTOXIN IS A PRESYNAPTIC NEUROTOXIN OF THE
VENOM. THE B CHAIN IS HOMOLOGOUS TO VENOM BASIC PROTEASE
INHIBITORS BUT HAS NO PROTEASE INHIBITOR ACTIVITY AND IS NON-
TOXIC.
CC -I- SUBUNIT: DIMER OF DISSIMILAR CHAINS LINKED BY A DISULFIDE BOND.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: Y12101; CAA72810.1; -
DR PIR: A01220; TIKF2.
DR HSSP: P00981; IDTK.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 1.
DR PRINTS: PR00759; BASICPTASE.
DR PRODOM: PD000222; Kunitz_BPTI; 1.
DR SMART: SM00131; Kunitz.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE: PS50279; BPTI_KUNITZ_2; 1.
KW Venom; Presynaptic neurotoxin; Signal.

FT SIGNAL 1 24
FT CHAIN 25 85
FT DOMAIN 31 81
FT DISULFID 31 81
FT DISULFID 40 64
FT DISULFID 56 77
FT DISULFID 79 79
FT ACT_SITE 41 42
FT CONFICT 44 44
FT CONFICT 65 70
FT CONFLICT 82 83
FT CONFLICT 82 83
SO SEQUENCE 85 AA; 9568 MW;
BETA-2 BUNGAROTOXIN B CHAIN.
BPTI/KUNITZ INHIBITOR.
BY SIMILARITY.
BY SIMILARITY.
INTERCHAIN (WITH AN A CHAIN) (POTENTIAL).
REACTIVE BOND (BY SIMILARITY).
MISSING (IN REF. 2).
NGNGNH -> DGDHGN (IN REF. 2).
LE -> EL (IN REF. 2).
FE95A59AF92BF2AA CRC64;

Query Match 3.7%; Score 6; DB 1; Length 85;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 84 LTPVSS 89
|||||
Db 19 LTPVSS 24

RESULT 14
ID YDGC_ECOLI STANDARD: PRT; 111 AA.

AC P52110; P77466;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydgc.
GN YDGC OR B1607 OR Z2608 OR ECS2313.
OS Escherichia coli, and
OC Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
ON NCBITaxid=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=K12;
RA Kuempel P.L.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G., Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.

RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alpha H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitagawa M., Kitagawa K.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takemoto K., Takuchi Y., Mada C., Yamamoto Y., Horichi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RN DNA Res. 3:363-377(1996).
RN [4]
RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.:
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.",
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 Rida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 Kahara S., Shiba T., Hattori M., Shinagawa H.;
 RT Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.*;
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP IDENTIFICATION.
 RA Rudd K.E.;
 RL Unpublished observations (JAN-1996).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Potential).
 CC -1- SIMILARITY: TO P.AERUGINOSA GLPM.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS
 CC IN POSITIONS 11 AND 38.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U41101; NOT ANNOTATED. CDS.
 DR EMBL: AE000256; AAC74679.1; -;
 DR EMBL: D90803; BAA15345.1; -;
 DR EMBL: D90804; BAA15345.1; -;
 DR EMBL: AE003584; AAC56594.1; -;
 DR EMBL: AP002558; BAB35736.1; -;
 DR EcoGene: EG13191; ydgc.
 KW Hypothetical protein; Transmembrane; Inner membrane;
 KM Complete proteome.
 FT TRANSMEM 27 47 POTENTIAL.
 FT TRANSMEM 59 79 POTENTIAL.
 FT TRANSMEM 88 108 POTENTIAL.
 SQ SEQUENCE 111 AA; 12323 MW; 8690061D033E5873 CRC64;

Query Match 3.7%; Score 6; DB 1; Length 111;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 92 GLVIKA 97
 ID 1111111
 DB 2 GLVIKA 7

RESULT 15
 REF3_YEAST STANDARD; PRT; 122 AA.
 ID REF3_YEAST
 AC P26755;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-OCT-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Replication factor A protein 3 (RF-A) (Replication protein A 13 kDa
 DE subunit).
 GN RFA3 OR YJL173C OR J0506.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycas.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 8-22; 29-46 AND 119-122.
 RC STRAIN=W303-1A;
 RX MEDLINE=91357474; PubMed=1885001;

RA Br11 S.J., Stillman B.;
 RT "Replication factor-A from Saccharomyces cerevisiae is encoded by
 RT three essential genes coordinately expressed at S phase.";
 RL Genes Dev. 5:1589-1600(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Loehning C., Mueller C., Freidel K., Clancy M.;
 RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP OVERHAUSER B., PIRAVANDI E., RINKE M., DOMLEY H.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDS SINGLE-STRANDED DNA. STIMULATES THE ACTIVITY
 CC OF A COGNATE STRAND EXCHANGE PROTEIN (SEPI). IT COOPERATES WITH
 CC T-AG AND DNA TOPOISOMERASE I TO UNWIND TEMPLATE DNA CONTAINING THE
 CC SIMIAN VIRUS 40 ORIGIN OF DNA REPLICATION.
 CC -1- SUBUNIT: HETEROTRIMER OF 69, 36, AND 13 KDA CHAINS. THE
 CC DNA-BINDING ACTIVITY MAY RESIDE EXCLUSIVELY ON THE 69 KDA SUBUNIT.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- PTM: The N-terminus is blocked.
 CC -1- SIMILARITY: LIMITED, TO MAMMALIAN RFA3/RPA3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X59750; CAA42422.1; -;
 DR EMBL: X56792; CAA40110.1; -;
 DR EMBL: Z49448; CAA89468.1; -;
 DR PIR: C37281; C37281.
 DR PIR: S17022; S17022.
 DR SGD: S0003709; RFA3.
 KW DNA replication; Nuclear protein.
 SQ SEQUENCE 122 AA; 13816 MW; 5EC5D6FB56BAC008 CRC64;

Query Match 3.7%; Score 6; DB 1; Length 122;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 63 SETPRV 68
 ID 1111111
 DB 3 SETPRV 8

RESULT 16
 ID NB5M_CHICK STANDARD; PRT; 133 AA.
 AC P48306;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative NADH-ubiquinone oxidoreductase B15 subunit (EC 1.6.5.3)
 DE (EC 1.6.99.3) (Complex I-B15) (CI-B15) (Hypothetical protein Walter)
 DE (GGHPW).
 GN NDUF84.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94040816; PubMed=7901127;
 RA Goldberg G.S., Kaczmarek W.;
 RT "A chicken genomic DNA fragment that hybridizes to the murine Hox-3.1
 RT subunit B15",
 RL gene 133:233-235(1993).
 RN [2]

```

RP SEQUENCE FROM N.A.
RX MEDLINE-93077061; PubMed-1359990;
RA Goldberg G.S., Kaczmarek W.;
RT "Sequence of a novel chicken genomic DNA fragment that hybridizes to
RL the murine Hox-3.1 homeobox.";
RT Gene 121:397-398(1992).
CC -1- FUNCTION: TRANSPORT OF ELECTRONS FROM NADH TO THE RESPIRATORY
CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC TO BE UBIQUINONE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinol = NAD(+) + ubiquinol.
CC -1- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane; matrix side
CC (by similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60778; CAA43193.1; ALT_SEQ.
DR EMBL; M84354; AAA70193.1; ALT_SEQ.
DR Hypothetical protein; Oxidoreductase; NAD: Ubiquinone: Mitochondrion.
SQ SEQUENCE 133 AA; 14938 MW; ED7E82948C97B352 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 133;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 87 VSSAGG 92
DB 117 VSSAGG 122

RESULT 17
YEDX_ECO57 STANDARD: PRT; 137 AA.
AC 08XB75;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transhyretin-like protein precursor.
GN YEDX OR Z3062 OR ECS2708.
OS Escherichia coli O157:H7.
OC Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE-21074935; PubMed-11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postaf J.G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RMD 0509952;
RX MEDLINE-21156331; PubMed-11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,
RA Kubura S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).

```

```

CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE TRANSTHYRETIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005418; AAG56983.1; -.
DR EMBL; AP002559; BAB36131.1; -.
DR InterPro; IPR000895; Transhyretin.
DR PROSITE; PS00768; TRANSTHYRETIN_1; 1.
DR PROSITE; PS00769; TRANSTHYRETIN_2; 1.
KW Transport; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 137 TRANSTHYRETIN-LIKE PROTEIN.
SQ SEQUENCE 137 AA; 15537 MW; 0A8D534B6A46B9F8 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 137;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 DVTYTL 147
DB 44 DVTYTL 49

RESULT 18
YEDX_ECOLI STANDARD: PRT; 137 AA.
AC P76341;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transhyretin-like protein precursor.
GN YEDX OR B1970.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP PARTIAL SEQUENCE OF N-TERMINUS.
RA Rudd K.E.;
RT Unpublished observations (Aug-1999).
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE TRANSTHYRETIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000288; AAC75036.1; -.
DR HSSP; P27731; TRP.
DR Ecogene; EG14046; yedX.
DR InterPro; IPR000895; Transhyretin.
DR Pfam; PF00576; Transhyretin; 1.

```

```

DR PRINTS: PRO0189; TRANSTHYRETIN.
DR ProDom: PD003457; Transthyretin; 1.
DR SMART: SM00095; TR_THY; 1.
DR PROSITE: PS00768; TRANSTHYRETIN_1; 1.
DR PROSITE: PS00769; TRANSTHYRETIN_2; 1.
KW Transport; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 23
FT CHAIN 24 137
SO SEQUENCE 137 AA; 15460 MW; B58A534051DDDC5E CRC64;

Query Match
Best Local Similarity 3.7%; Score 6; DB 1; Length 137;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 DVTYTL 147
DB 44 DVTYTL 49

RESULT 19
RS9_PYRAE STANDARD; PRT; 142 AA.
ID RS9_PYRAE
AC ORFYO0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 30S ribosomal protein S9P.
GN RPS9P OR PAE0674.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX PubMed=11792869;
RA Ritz-Gibson S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT 'genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum';
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC -1- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE009779; AAL62943.1; ALT_INIT.
DR InterPro: IPR000754; Ribosomal_S9.
DR Pfam: PF00380; Ribosomal_S9; 1.
DR ProDom: PD001627; Ribosomal_S9; 1.
DR PROSITE: PS00360; RIBOSOMAL_S9; 1.
KW Ribosomal protein; Complete proteome.
SO SEQUENCE 142 AA; 15930 MW; 8D31075F2E4CF842 CRC64;

Query Match
Best Local Similarity 3.7%; Score 6; DB 1; Length 142;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 ETPRVV 69
DB 9 ETPRVV 14

RESULT 20
H2A2_PEA STANDARD; PRT; 149 AA.
ID H2A2_PEA
AC P40281;
DT 01-FEB-1995 (Rel. 31, Created)

```

```

DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone H2A.
OS Pisum sativum (Garden pea).
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
CC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alaska;
RX MEDLINE=96046745; PubMed=7579177;
RA Devitt M.L., Stafrum J.P.;
RT "Cell cycle regulation during growth-dormancy cycles in pea axillary
RT buds.";
RL Plant Mol. Biol. 29:255-265(1995).
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H2A FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U10041; AAA86947.1; -.
DR InterPro: IPR002119; Histone_H2A.
DR InterPro: IPR004822; Histone_core.
DR Pfam: PF00125; histone; 1.
DR PRINTS: PR00620; HISTONE_H2A.
DR ProDom: PD000522; Histone_H2A; 1.
DR SMART: SM00414; H2A; 1.
DR PROSITE: PS00046; HISTONE_H2A; 1.
KW Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding;
KW Multigene family; Acetylation.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT SEQUENCE 149 AA; 15704 MW; 5ED29866408BC05 CRC64;

Query Match
Best Local Similarity 3.7%; Score 6; DB 1; Length 149;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 AYGGVL 43
DB 112 AYGGVL 117

RESULT 21
H2A1_PEA STANDARD; PRT; 150 AA.
ID H2A1_PEA
AC P25470;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone H2A.
OS Pisum sativum (Garden pea).
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
CC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=93044503; PubMed=1841722;
RA Konig A.J., Tanimoto E.Y., Kiehn K., Rost T., Comai L.;
RT "Cell-specific expression of plant histone H2A genes.";
RL Plant Cell 3:657-665(1991).
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.

```

```

CC -1- SIMILARITY: BELONGS TO THE HISTONE H2A FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M64838; -; NOT_ANNOTATED_CDS.
CC PIR: J01183; J01183.
CC InterPro: IPR002119; Histone_H2A.
CC InterPro: IPR004822; Histone_core.
CC Pfam: PF00125; histone; 1.
CC PRINTS: PR00620; HISTONEH2A.
CC ProDom: PD000522; Histone_H2A; 1.
CC SMART: SM00414; H2A; 1.
CC PROSITE: PS00046; HISTONE_H2A; 1.
CC Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding;
CC Multigene family; Acetylation.
CC MOD_RES 1 ACETYLATION (BY SIMILARITY).
CC SEQUENCE 150 AA; 15879 MW; 0F72BAFE5265F230 CRC64;
SQ
Query Match 3.7%; Score 6; DB 1; Length 150;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 38 AYGVL 43
|11111|
DB 112 AYGVL 117

RESULT 22
NAPB_RHOSH STANDARD; PRT; 154 AA.
ID NAPB_RHOSH
AC 053177;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Diheme cytochrome c napb precursor.
GN NAPB.
OS Rhodobacter sphaeroides (Rhodospirillum rubrum sphaeroides).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
CC Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX MEDLINE=96332666; PubMed=8730872;
RA Reyes F., Roldan M.D., Klipp W., Castillo F., Moreno-Vivian C.;
RT "Isolation of periplasmic nitrate reductase genes from Rhodobacter
RT sphaeroides DSM 158: structural and functional differences among
RT prokaryotic nitrate reductases."
RT Mol. Microbiol. 19:1307-1318(1996).
CC -1- FUNCTION: SMALL SUBUNIT OF THE PERIPLASMIC NITRATE REDUCTASE
CC (NAPB). ONLY EXPRESSED AT HIGH LEVELS DURING AEROBIC GROWTH. NAPAB
CC COMPLEX RECEIVES ELECTRONS FROM THE MEMBRANE-ANCHORED TETRAHEME
CC NADH PROTEIN, THUS ALLOWING ELECTRON FLOW BETWEEN MEMBRANE AND
CC PERIPLASM. ESSENTIAL FUNCTION FOR NITRATE ASSIMILATION AND MAY
CC HAVE A ROLE IN ANAEROBIC METABOLISM.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC SUBUNIT AND A CYTOCHROME C (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: INDUCED BY NITRATE. NOT REPRESSED BY AMMONIUM OR
CC OXYGEN.
CC -1- PTM: BINDS TWO HEME GROUPS PER MOLECULE (BY SIMILARITY).
CC -1- SIMILARITY: STRONG, TO OTHER BACTERIAL NAPB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z46806; -; NOT_ANNOTATED_CDS.
CC InterPro: IPR000345; Cyt_c_heme_bind.
CC PROSITE: PS00190; CYTOCHROME_C; 2.
CC Electron transport; Heme; Periplasmic; Signal.
CC SIGNAL 1 24
CC CHAIN 25 154
CC MOD_RES 25 25
CC BINDING 82 82
CC BINDING 85 85
CC METAL 86 86
CC FT BINDING 122 122
CC FT BINDING 125 125
CC FT METAL 126 126
CC SEQUENCE 154 AA; 16908 MW; 63CE740A5864977 CRC64;
SQ
Query Match 3.7%; Score 6; DB 1; Length 154;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 24 PETTD 29
|11111|
DB 46 PETTD 51

RESULT 23
LSPA_STACA STANDARD; PRT; 159 AA.
ID LSPA_STACA
AC Q59835;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipoprotein signal peptidase (EC 3.4.23.36) (Prolipoprotein signal
DE peptidase) (Signal peptidase II) (Spase II).
GN LSPA OR LSP.
OS Staphylococcus carnosus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1281;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FM300;
RX MEDLINE=95247010; PubMed=7729667;
RA Wilke C., Goetz F.;
RT "Cloning and nucleotide sequence of the signal peptidase II (lsp)-gene
RT from Staphylococcus carnosus."
RT FEMS Microbiol. Lett. 126:233-240(1995).
CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY CATALYZES THE REMOVAL OF
CC SIGNAL PEPTIDES FROM PROLIPOPROTEINS.
CC -1- CATALYTIC ACTIVITY: Cleavage of N-terminal leader sequences from
CC membrane prolipoproteins. Hydrolyzes xaa-xbb-xcc-1-cys, in which
CC xaa is hydrophobic (preferably leu), xbb is often ser or ala, xcc
CC is often gly or ala, and the cys is alkylated on sulfur with a
CC diacylglyceryl group.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X78084; CAA54991.1; -.
CC MEPROPS; A08.001; -.
CC InterPro: IPR001872; SigPTase_A8.
CC Pfam: PF01252; Peptidase_A8; 1.

```

```

DR PRINTS; PRO0781; LIPOSICPTASE.
DR PRODOM; PD004304; S1PTase_A8; 1.
DR TIGRFAMS; TIGR00077; lspa; 1.
DR PROSITE; PS00855; SPASE_II; 1.
KM Hydrolase; Aspartyl protease; Transmembrane.
FT DOMAIN 1 3 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 4 24 POTENTIAL.
FT TRANSMEM 25 63 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 64 84 POTENTIAL.
FT TRANSMEM 85 87 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 88 108 POTENTIAL.
FT TRANSMEM 109 130 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 131 131 POTENTIAL.
FT DOMAIN 132 159 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 109 109 BY SIMILARITY.
FT ACT_SITE 136 136 BY SIMILARITY.
SQ SEQUENCE 159 AA; 17630 MW; F9DF08E22811FBD CRC64;

Query Match 3.7%; Score 6; DB 1; Length 159;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 IAVLIL 106
DB 15 IAVLIL 20

RESULT 24
RL6_BACSU STANDARD; PRT; 178 AA.
AC P46988;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L6 (BL10).
GN RPL6.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168; Warburg;
RX MEDLINE=96186897; PubMed=8635744;
RA Suh J.W., Boylan S.A., Oh S.H., Price C.W.;
RT "Genetic and transcriptional organization of the Bacillus subtilis
    spc-alpha region."
RL Gene 169:17-23(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124188; PubMed=8695901;
RA Yasunoto K., Liu H., Jeong S.M., Ohashi Y., Kakimura S.,
    Tanaka K., Kawamura F., Yoshikawa H., Takahashi H.;
RT "Sequence analysis of a 50 kb region between spoOH and trnH on the
    Bacillus subtilis chromosome."
RL Microbiology 142:3039-3046(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
    Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
    Bortiss R., Bourcier L., Brans A., Braun M., Brynneil S.C., Bron S.,
    Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
    Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
    Denizot F., Devine K.M., Dusterhoft A., Entrlich S.D., Emmertson P.T.,
    Ertlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
    Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
    Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
    Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
    Hilbert H., Holsappel S., Hosono S., Huilo M.F., Itaya M., Jones L.,
    Joris B., Karamata D., Kasahara Y., Kletter-Bianchard M., Klein C.,
    Kobayashi Y., Koelter P., Konigstein G., Krogh S., Kumano M.,

```

```

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
    Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
    Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
    Noone D., O'Reilly M., Ogawa K., Ogihara K., Oudega B., Park S.H.,
    Parro V., Pohl T.M., Portetelle D., Potworlik S., Prescott A.M.,
    Prescan E., Puig P., Purnelle D., Rapoport G., Rey M., Reynolds S.,
    Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadleir Y.,
    Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,
    Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
    Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
    Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
    Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
    Viari A., Wambutt R., Wedler E., Wedler H., Weitzengger T.,
    Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
    Yoshida K., Yoshikawa H.F., Zunshtein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
    subtilis."
RL Nature 390:249-256(1997).
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND IS
    LOCATED AT THE AMINOACYL-TRNA BINDING SITE OF THE
    PEPTIDYLTRANSFERASE CENTER.
CC -1- SIMILARITY: BELONGS TO THE L6P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL Outstation
    at the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
CC
DR EMBL; I47971; AAB06814.1; -
DR EMBL; D64125; BAA10981.1; -
DR EMBL; Z69104; CAB11907.1; -
DR HSSP; P02391; RL6.
DR Subtilisin; BG11408; rPLF.
DR InterPro; IPR00702; Ribosomal_L6.
DR InterPro; IPR002358; Ribosomal_L6_1.
DR Pfam; PF00347; Ribosomal_L6; 1.
DR PRINTS; PRO0059; RIBOSOMAL_L6.
DR PRODOM; PD002236; RIBOSOMAL_L6; 1.
DR PROSITE; PS00525; RIBOSOMAL_L6_1; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
FT INIT_MET 0 BY SIMILARITY.
SQ SEQUENCE 178 AA; 19378 MW; A560F56DC66CFDC CRC64;

Query Match 3.7%; Score 6; DB 1; Length 178;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 DVTVTL 147
DB 13 DVTVTL 18

RESULT 25
PTH_CHLMU STANDARD; PRT; 179 AA.
AC P49607;
DT 01-FEB-1996 (Rel. 33, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peptidyl-trna hydrolase (EC 3.1.1.29) (PTH).
GN PTH OR TC0183.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN;
RX MEDLINE=96186910; PubMed=8635758;
RA de la Vega F.M., Galindo J.M., Old I.G., Guarneros G.;
RT "Microbial genes homologous to the peptidyl-trna hydrolase-encoding

```

```

RT gene of Escherichia coli.;
RL Gene 169:97-100(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MOPn / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
  White O., Hickey E.K., Peterson J., Uitterback T., Berry K.,
  Bass S., Liner K., Weidman J., Khouri H., Civen B., Bowman C.,
  Dodoson R., Gwin M., Nelson W., Deboy R., Koloney J., McClarty G.,
  Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
  pneumoniae A339."
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE OF 27-194 FROM N.A.
RC STRAIN-MOPn;
RX MEDLINE=91072247; PubMed=2254267;
RA Engel J.N., Pollack J., Petara E., Ganem D.;
RT "Heat shock response of murine Chlamydia trachomatis.";
RL J. Bacteriol. 172:6959-6972(1990).
CC -1- FUNCTION: THE NATURAL SUBSTRATE FOR THIS ENZYME MAY BE PEPTIDYL-
  TNRAS WHICH DROP OFF THE RIBOSOME DURING PROTEIN SYNTHESIS
  (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N-substituted aminoacyl-tRNA + H(2)O = N-
  substituted amino acid + tRNA.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PRN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U31570; AAB06184.1; -
DR EMBL; AE002285; AAF39057.1; -
DR EMBL; M62820; AAA23133.1; -
DR PIR; C37840; C37840.
DR HSSP; P23932; 2PTH.
DR TIGR; T20183; -
DR InterPro: IPR001328; Pept_tRNA_hydro.
DR Pfam; PF01195; Pept_tRNA_hydro; 1.
DR ProDom; PD005324; Pept_tRNA_hydro; 1.
DR TIGRFAMs; TIGR00447; plh; 1.
DR PROSITE; PS01195; PEPT_TRNA_HYDROL_1; 1.
DR PROSITE; PS01196; PEPT_TRNA_HYDROL_2; 1.
KW Hydrolyase; Complete proteome.
FT CONFLICT 146 146 Y->G (IN REF. 1 AND 3).
FT CONFLICT 161 161 S->G (IN REF. 1 AND 3).
FT CONFLICT 178 178 GF->AFNKGIVCSLFEKRS (IN REF. 1 AND
  3).
SQ SEQUENCE 179 AA; 19915 MW; 5D682B5269C0564B CRC64;

Query Match 3.7%; Score 6; DB 1; Length 179;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 42 VLSNFS 47
DB 150 VLSNFS 155

```

```

DE Holliday junction DNA helicase ruva.
GN RUVA OR FM1104.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
ON NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
  Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
  Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
  Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
  Feinstein M., Kyrpides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
  nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -1- FUNCTION: The ruva-ruvb complex in the presence of ATP renatures
  cruciform structure in supercoiled DNA with palindromic sequence,
  indicating that it may promote strand exchange reactions in
  homologous recombination. RuvaB is an helicase that mediates the
  Holliday junction migration by localized denaturation and
  reannealing. Ruva stimulates, in the presence of DNA, the weak
  ATPase activity of ruvaB (By similarity).
CC -1- SUBUNIT: Forms a complex with ruvb.
CC -1- SIMILARITY: BELONGS TO THE RUVA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE010615; AAL95300.1; -
DR DNA repair; SOS response; DNA-binding; DNA recombination; Helicase;
  Complete proteome.
SQ SEQUENCE 194 AA; 22579 MW; DA0381C5FB8B0F3 CRC64;

Query Match 3.7%; Score 6; DB 1; Length 194;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 42 VLSNFS 47
DB 88 VLSNFS 93

```

```

RESULT 27
RL3_BUCAI
ID ID R13_BUCAI STANDARD: PRT: 209 AA.
AC P57591;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L3.
GN RPIC OR BU524.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
  symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
  Buchnera sp. Aps."
RL Nature 407:81-86(2000).
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND MAY
  PARTICIPATE IN THE FORMATION OF THE PEPTIDYLTRANSFERASE CENTER OF
  THE RIBOSOME (BY SIMILARITY).

```

CC -1- SIMILARITY: BELONGS TO THE L3P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB001119; BAB13217.1; -
 DR InterPro: IPR000397; Ribosomal_L3.
 DR Pfam: PF00297; Ribosomal_L3; 1.
 DR ProDom: PD001374; Ribosomal_L3; 1.
 DR PROSITE: PS00474; RIBOSOMAL_L3; 1.
 KM Ribosomal protein; rRNA-binding; Complete proteome.
 SQ SEQUENCE 209 AA; 22841 MW; CF0D7F60C828C76F CRC64;

Query Match 3.7%; Score 6; DB 1; Length 209;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 FSGTVK 51
 |||||
 Db 118 FSGTVK 123

RESULT 28
 MUC3_HUMAN
 ID MUC3_HUMAN STANDARD; PRT; 213 AA.
 AC 002505; Q02506;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Mucin 3A (Intestinal mucin 3A) (Fragments).
 GN MUC3A OR MUC3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Small intestine;
 RX MEDLINE=90365738; PubMed=2393399;
 RA Gum J.R. Jr., Hicks J.W., Swallow D.M., Lagace R.L., Byrd J.C.,
 RA Lamport D.T.A., Siddiki B., Kim Y.S.;
 RT "Molecular cloning of cDNAs derived from a novel human intestinal
 RT mucin gene.";
 RL Biochem. Biophys. Res. Commun. 171:407-415(1990).
 CC -1- FUNCTION: MAJOR GLYCOPROTEIN COMPONENT OF A VARIETY OF MUCUS GELS.
 CC THOUGHT TO PROVIDE A PROTECTIVE, LUBRICATING BARRIER AGAINST
 CC PARTICLES AND INFECTIOUS AGENTS AT MUCOSAL SURFACES.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: BROAD SPECIFICITY; SMALL INTESTINE, COLON AND
 CC COLONIC TUMORS.
 CC -1- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M55406; AAA63773.1; -
 DR EMBL: M55405; AAA63772.1; -
 DR PIR: A35690; A35690.
 DR PIR: B35690; B35690.
 DR Genew: HGNC:7513; MUC3A.
 DR MIM: 158371; -
 KM Glycoprotein; Repeat.
 FT NON_TER 1

FT DOMAIN <1 >213 REPEATS.
 FT REPEAT <1 9
 FT REPEAT 10 26
 FT REPEAT 27 49
 FT REPEAT 27 49
 FT REPEAT 50 66
 FT REPEAT 67 83
 FT REPEAT 84 100
 FT REPEAT 101 117
 FT REPEAT 118 >128
 FT NON_CONS 128 129
 FT REPEAT <129 145
 FT REPEAT 146 162
 FT REPEAT 163 179
 FT REPEAT 180 196
 FT REPEAT 197 >213
 FT NON_TER 213
 SQ SEQUENCE 213 AA; 22042 MW; 064EC7A237D786E CRC64;

Query Match 3.7%; Score 6; DB 1; Length 213;
 Best Local Similarity 100.0%; Pred. No. 16+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 TTSETP 66
 |||||
 Db 3 TTSETP 8

RESULT 29
 RPB5_KLUMA
 ID RPB5_KLUMA STANDARD; PRT; 215 AA.
 AC Q9P4B9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA-directed RNA polymerase II subunit 5).
 GN RPB5.
 OS Kluveromyces marxianus (Yeast) (Kluveromyces fragilis).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluveromyces.
 OX NCBI_TaxID=4911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12424;
 RX MEDLINE=20432119; PubMed=10974568;
 RA Ladrriere J.-M., Georis I., Guerin M., Vandenhaute J.;
 RA "Kluveromyces marxianus exhibits an ancestral Saccharomyces
 RT cerevisiae genome organization downstream of ADH2.";
 RL Gene 255:83-91(2000).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INFO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 15
 CC DIFFERENT POLYPEPTIDES. THIS SUBUNIT IS SHARED BY ALL 3 YEAST RNA
 CC POLYMERASES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
 CC III FOR 5S AND TRNA GENES.
 CC -1- SIMILARITY: BELONGS TO THE ARCHAEABACTERIA RPOH / EUKARYOTIC RPB5
 CC RNA POLYMERASE SUBUNIT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF225206; AAF91238.1; -
DR HSSP; P20434; 1DZF.
DR InterPro: IPR000783; RNA_POL_H_23KD.
DR Pfam; PF01191; RNA_POL_H; 1.23KD; 1.
DR ProDom; PD005155; RNA_POL_H_23KD; 1.
DR PROSITE; PS01110; RNA_POL_H_23KD; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription;
Nuclear protein.
SO SEQUENCE 215 AA; 24954 MW; 10D7EA1754B3ACF0 CRC64;

Query Match 3.7%; Score 6; DB 1; Length 215;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 PVALYL 84
DB 183 PVALYL 188

RESULT 30
RPB5_YEAST STANDARD: PRT: 215 AA.
AC P20434; Q02121;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerases I, II, and III 27 kDa polypeptide
DE (EC 2.7.7.6) (AEC27).
GN RPB5 OR YBR154C OR YBR1204.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=90245736; PubMed=2186966;
RX Mochlik N.A., Liao S.-M., Kolodziej P.A., Young R.A.;
RT "Subunits shared by eukaryotic nuclear RNA polymerases.";
RL Genes Dev. 4:313-323(1990).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=S288C;
RX MEDLINE=93255907; PubMed=8488729;
RA Baur A., Schaaff-Gerstenschlaeger I., Boles E., Miosga T., Rose M.,
RL Zimmermann F.K.;
RT "Sequence of a 4.8 kb fragment of Saccharomyces cerevisiae chromosome
RT II including three essential open reading frames.";
RL Yeast 9:289-293(1993).
[3]
RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RP MEDLINE=20300896; PubMed=10841537;
RA Todone F., Weinzierl R.O.J., Brick P., Onesti S.;
RT "Crystal structure of RPB5, a universal eukaryotic RNA polymerase
RT subunit and transcription factor interaction target.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:6306-6310(2000).
-1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
SUBSTRATES.
-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
[RNA](N).
-1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 15
DIFFERENT POLYPEPTIDES. THIS SUBUNIT IS SHARED BY ALL 3 YEAST RNA
POLYMERASES.
-1- SUBCELLULAR LOCATION: Nuclear.
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
III FOR 5S AND TRNA GENES.
CC -1- SIMILARITY: BELONGS TO THE ARCHAEABACTERIA RPOH / EUKARYOTIC RPB5
RNA POLYMERASE SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53287; CAA37381.1; -
DR EMBL; X71329; CAA50472.1; -
DR EMBL; S59774; AAC60556.1; -
DR EMBL; Z36023; CAA85113.1; -
DR PIR; A34588; A34588.
DR PDB; 1DZF; 09-JUN-00.
DR SCD; S0000358; RPB5.
DR InterPro: IPR000783; RNA_POL_H_23KD.
DR Pfam; PF01191; RNA_POL_H; 1.23KD; 1.
DR ProDom; PD005155; RNA_POL_H_23KD; 1.
DR PROSITE; PS01110; RNA_POL_H_23KD; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription;
Nuclear protein; 3D-structure.
FT CONFLICT 37 37 L -> W (IN REF. 2; AAC60556).
FT SEQUENCE 215 AA; 25079 MW; 2A15F71AD69D829 CRC64;

Query Match 3.7%; Score 6; DB 1; Length 215;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 PVALYL 84
DB 183 PVALYL 188

RESULT 31
ALKB_SALTY STANDARD: PRT: 216 AA.
AC P37462;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alkylated DNA repair protein alkb.
GN ALKB OR STM2264.
OS Salmoneilla typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmoneilla.
OX NCBI_TaxID=602;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=LT2 / SGC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard L., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmoneilla enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
[2]
RN SEQUENCE OF 1-24 FROM N.A.
RP MEDLINE=91267928; PubMed=1904855;
RX Hakura A., Morimoto K., Sofuni T., Nohmi T.;
RT "Cloning and characterization of the Salmoneilla typhimurium ada gene,
RT which encodes O6-methylguanine-DNA methyltransferase.";
RL J. Bacteriol. 173:3663-3672(1991).
-1- FUNCTION: PROBABLY INVOLVED IN THE REPAIR OF ALKYLATED DNA.
-1- SIMILARITY: BELONGS TO THE ALKB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----


```

CC -----
DR EMBL; AE008801; AAL2116.1; -.
DR D90221; NOT_ANNOTATED_CDS.
DR PIR; B39433; B39433.
DR SEVGENE; SG10006; alkb.
KM DNA repair; Complete proteome.
FT CONFLICT 19 A -> R (IN REF. 2).
SQ SEQUENCE 216 AA; 23787 MW; 8F320C7B9CABFAB6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 216;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 TDRKWP 79
   |||||
Db 85 TDRKWP 90

RESULT 32
CAT_PROMI STANDARD; PRT; 217 AA.
ID CAT_PROMI
AC P07641;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Chloramphenicol acetyltransferase (EC 2.3.1.28).
GN CAT.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID=584;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=86008040; Pubmed=3900035;
RA Charles I.G., Keyte J.W., Shaw W.V.;
RT "Nucleotide sequence analysis of the cat gene of Proteus mirabilis:
RL J. Bacteriol. 164:123-129(1985).
CC -1- FUNCTION: THIS ENZYME IS AN EFFECTOR OF CHLORAMPHENICOL RESISTANCE
IN BACTERIA.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + chloramphenicol = CoA +
chloramphenicol 3-acetate.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SIMILARITY: BELONGS TO THE CHLORAMPHENICOL ACETYLTRANSFERASE
FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M1587; AAA25655.1; -.
DR PIR; A24651; A24651.
DR HSSP; P00484; 10CA.
DR InterPro: IPR001707; CAT.
DR Pfam: PF003102; CAT.1.
DR PRODOM; PD002660; CAT.1.
DR PROSITE; PS00100; CAT.1.
KW Antibiotic resistance; Transferase; Acyltransferase.
FT ACT_SITE 193 193 BY SIMILARITY.
SQ SEQUENCE 217 AA; 25313 MW; 08ABBA43F9FC41C2 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 217;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 LYYDLS 14
   |||||
Db 9 LYYDLS 14

```

```

RESULT 33
R10A_SPOFR STANDARD; PRT; 217 AA.
ID R10A_SPOFR
AC Q963B6;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L10a.
GN RPL10A.
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptera; Noctuidae; Noctuidae; Amphipyrinae; Spodoptera.
OX NCBI_TaxID=7108;
RN [1]
RP SEQUENCE FROM N.A.
RA Landais I., Ogilastro M., Mila K., Nohata J., Lopez-Ferber M.,
RA Duonor-Cerutti M., Fournier P., Devauchelle G.;
RT "Full-length ribosomal protein sequence from an EST library of
RL Spodoptera frugiperda cells (Sf9)."
CC Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE LIP FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF391092; AAK76990.1; -.
DR InterPro: IPR002143; Ribosomal_L1.
DR Pfam: PF00687; Ribosomal_L1; 1.
DR PROSITE; PS01199; RIBOSOMAL_L1; FALSE_NEG.
SQ SEQUENCE 217 AA; 24518 MW; 9E284AAE1FPA082C CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 217;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 FSGTIV 51
   |||||
Db 49 FSGTIV 54

RESULT 34
VT11_YEAST STANDARD; PRT; 217 AA.
ID VT11_YEAST
AC 004358;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vesicle transport v-SNARE protein VT11.
GN VT11 OR YMR197C OR YMR646.10C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97344273; Pubmed=9199167;
RA Fischer von Mollard G., Notwehr S.F., Stevens T.H.;
RT "The yeast v-SNARE Vti1p mediates two vesicle transport pathways
RL through interactions with the t-SNAREs Sed5p and Pcp12p."
RN J. Cell Biol. 137:1511-1524(1997).
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;

```

Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.

[3] CHARACTERIZATION. PubMed-9398683; MEDLINE-98062344; McNew J.A., Waters M.G.; Lupashin V.V., Pokrovskaya I.D., McNew J.A., Waters M.G.; "Characterization of a novel yeast SNARE protein implicated in Golgi retrograde traffic."; Mol. Biol. Cell 8:2659-2676(1997).

[4] BINDING TO T-SNARES TUG1 AND TUG2. MEDLINE-98090455; PubMed-9427746; Holthuis J.C., Nichols B.J., Dhruvakumar S., Pelham H.R.; "Two syntaxin homologues in the TGN/endosomal system of yeast."; EMBO J. 17:113-126(1998).

[5] INVOLVEMENT IN HOMOTYPIC VACUOLAR FUSION. MEDLINE-99315912; PubMed-10385523; Ungermann C., Fischer von Mollard G., Jensen O.N., Margolis N., Stevens T.H., Wickner W.; "Three v-SNARES and two t-SNARES, present in a pentameric cis-SNARE complex on isolated vacuoles, are essential for homotypic fusion."; J. Cell Biol. 145:1435-1442(1999).

[6] INVOLVEMENT IN MULTIPLE TRAFFICKING STEPS. MEDLINE-99287719; PubMed-10359592; Fischer von Mollard G., Stevens T.H.; "The Saccharomyces cerevisiae v-SNARE Vti1p is required for multiple membrane transport pathways to the vacuole."; Mol. Biol. Cell 10:1719-1732(1999).

[7] FUNCTION: V-SNARE ESSENTIAL FOR YEAST CELL VIABILITY WHICH MEDIATES MULTIPLE VESICLE TRANSPORT PATHWAYS THROUGH INTERACTIONS WITH T-SNARES ON THE TARGET MEMBRANE OF DIFFERENT ORGANELLES. THESE INTERACTIONS ARE PROPOSED TO MEDIATE ASPECTS OF THE SPECIFICITY OF VESICLE TRAFFICKING AND TO PROMOTE FUSION OF THE LIPID BILAYERS.

[8] SUBUNIT: FORMS A SNARE COMPLEX WITH THE CIS-GOLGI T-SNARE SED5 AND V-SNARES SPT1 AND YTK6, WHICH IS INVOLVED IN RETROGRADE TRAFFIC TO THE CIS-GOLGI COMPARTMENT. INTERACTS WITH THE PREVACUOLAR T-SNARE PEP12 IN TRAFFIC FROM THE TRANS-GOLGI NETWORK (TGN) TO THE PREVACUOLAR COMPARTMENT (PVC). INTERACTS WITH THE VACUOLAR T-SNARE VMA3 IN BIOSYNTHETIC TRANSPORT PATHWAYS TO THE VACUOLE AND IN HOMOTYPIC VACUOLAR FUSION. BINDS ALSO TO THE EARLY ENDOSOME T-SNARE TUG1 AND THE TGN T-SNARE TUG2.

[9] SUBCELLULAR LOCATION: A SMALL PORTION IS LOCALIZED IN THE GOLGI APPARATUS. THE MAJORITY IS LOCALIZED IN THE PVC.

[10] MISCELLANEOUS: TEMPERATURE-SENSITIVE (TS) MUTANTS Vti1-1 AND Vti1-2 EXHIBIT DEFECTS IN TGN TO PVC TRANSPORT AT THE NONPERMISSIVE TEMPERATURE. Vti1-11 MUTANT DISPLAY A BLOCK IN TRAFFIC TO THE PVC AND AN ADDITIONAL DEFECT IN RETROGRADE TRAFFIC TO THE CIS-GOLGI. Vti1-12 MUTANT HAS A CONSTITUTIVE BLOCK OF TRAFFIC FROM THE LATE GOLGI TO THE VACUOLE AND A TS BLOCK TO THE CIS-GOLGI COMPARTMENT.

[11] SIMILARITY: BELONGS TO THE VTI1 FAMILY.

[12] SIMILARITY: CONTAINS 1 T-SNARE COILED-COIL HOMOLOGY DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AF006074; AAC49745.1; -

DR EMBL: Z47815; CA87819.1; -

DR SGP: S0004810; VTI1.

DR InterPro: IPR000727; T-SNARE.

DR SMART: SM00397; T-SNARE; 1.

DR PROSITE: PS0192; T-SNARE; 1.

KW Transmembrane; Transport; Protein transport; Coiled coil.

FT DOMAIN 124 186 T-SNARE COILED-COIL HOMOLOGY.

FT TRANSMEM 195 215 POTENTIAL.

FT VARIANT 8 8 Y -> R (IN MUTANT VTI1-11).

FT VARIANT 20 20 K -> R (IN MUTANT VTI1-11).

FT VARIANT 40 40 H -> R (IN MUTANT VTI1-11).

FT VARIANT 61 61 N -> S (IN MUTANT VTI1-11).

FT VARIANT 73 73 K -> R (IN MUTANT VTI1-11).

FT VARIANT 84 84 Q -> R (IN MUTANT VTI1-11).

FT VARIANT 130 130 A -> S (IN MUTANT VTI1-12).

FT VARIANT 141 141 E -> G (IN MUTANT VTI1-11).

FT VARIANT 145 145 E -> G (IN MUTANT VTI1-11).

FT VARIANT 145 145 E -> K (IN MUTANT VTI1-11).

FT VARIANT 148 148 G -> R (IN MUTANT VTI1-11).

FT VARIANT 155 155 L -> F (IN MUTANT VTI1-11).

FT VARIANT 158 158 Q -> R (IN MUTANT VTI1-12).

FT VARIANT 158 158 Q -> R (IN MUTANT VTI1-12).

SO SEQUENCE 217 AA; 24668 MW; E5B4DA3B40FAD67 CRC64;

Query Match 3.7%; Score 6; DB 1; Length 217;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 IAVLL 106

IIIIII

Db 201 IAVLL 206

RESULT 35

GCH1_HAEIN STANDARD; PRT; 218 AA.

ID GCH1_HAEIN

AC P43866;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE GTP cyclohydrolase I (EC 3.5.4.16) (GTP-CH-I).

GN FOLE OR H1447.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

CC Haemophilus.

CC NCBI_TaxID=727;

OX [1]

RN SEQUENCE FROM N.A.

RP STRAIN-Rd / KW20 / ATCC 51907;

RC MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kienlevey A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;

RA "Whole-genome random sequencing and assembly of Haemophilus influenzae Venter J.C.;

RT Rd.;"

RL Science 269:496-512(1995).

CC -I- CATALYTIC ACTIVITY: GTP + 2 H(2)O = formate + 2-amino-4-hydroxy-6-(gamma-thio-1,2,3-trihydroxypropyl)dihydropteridine triphosphate.

CC -I- PATHWAY: Tetrahydrofolate biosynthesis; first step.

CC -I- SUBUNIT: HOMOPOLYMER (BY SIMILARITY).

CC -I- SIMILARITY: BELONGS TO THE GTP CYCLOHYDROLASE I FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U32823; AAC23097.1; -

DR HSSP: P27511; IABR.

DR TIGR: H1447; -

DR InterPro: IPR001474; GTP_cyclohydrol.

DR Pfam: PF01227; GTP_cyclohydrol; 1.

DR ProDom: PD003330; GTP_cyclohydrol; 1.

DR TIGRFAMS: TIGR00063; FOLE; 1.

DR PROSITE: PS00859; GTP-CYCLOHYDROL_1_1; 1.

DR PROSITE: P500860; GTP_CYCLOHYDROL_1.2; 1.
 KW One-carbon metabolism: Hydrolase; Complete proteome.
 FT DISULFID 109 180 BY SIMILARITY
 SQ SEQUENCE 218 AA; 24945 MW; BCC60F003815BD38 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 6; DB 1; Length 218;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 37 SAGGV 42
 |||||
 DB 196 SAGGV 201

RESULT 36
 GNP_ECOLI STANDARD: PRT; 219 AA.
 AC P10345; P76825; (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-MAR-1989 (Rel. 10, Last annotation update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glutamine transport system permease protein glnP.
 GN GNP OR B0810 OR Z1032 OR ECS0888.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=87115160; PubMed=3027504;
 RA Nono T., Saito T., Hong J.;
 RT "Cloning and complete nucleotide sequence of the Escherichia coli
 RL glutamine permease operon (glnHP)";
 RN Mol. Gen. Genet. 205:260-269(1986).
 [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.D.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 [13]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizubuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.;
 RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
 RL corresponding to the 12.7-28.0 min region on the linkage map";
 RN DNA Res. 3:137-155(1996).
 [14]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boulton A., Shao Y., Miller L.,
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamovasis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blatter F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RN Nature 409:529-533(2001).
 [15]
 RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RL O157:H7 and genomic comparison with a laboratory strain K-12";
 RN DNA Res. 8:11-22(2001).
 [1]
 CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
 CC FOR GLUTAMINE; PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE
 CC SUBSTRATE ACROSS THE MEMBRANE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -1- INDUCTION: BY LACK OF GLUTAMINE.
 CC -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
 CC SYSTEM PERMEASE FAMILY. HISMO SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X14180; CA33383.1; -;
 DR EMBL: AE000183; AAC73897.1; -;
 DR EMBL: D90718; BAA35482.1; -;
 DR EMBL: D90719; BAA35492.1; -;
 DR EMBL: AE005262; AAG55182.1; -;
 DR EMBL: AP002553; BAR34311.1; -;
 DR PIR: S03182; ORECP.
 DR Ecogene; EG10388; glnP.
 DR InterPro: IPR00515; BPD_transp.
 DR Pfam: PF00528; BPD_transp. 1.
 DR PROSITE: P500402; BPD_TRANSPO. INN_MEMBR. 1.
 DR Transprot; Amino-acid transport; Transmembrane; Inner membrane;
 KW Complete proteome.
 FT TRANSMEM 23 43
 FT TRANSMEM 54 74 POTENTIAL.
 FT TRANSMEM 89 109 POTENTIAL.
 FT TRANSMEM 151 171 POTENTIAL.
 FT TRANSMEM 188 208 POTENTIAL.
 SQ SEQUENCE 219 AA; 24364 MW; 20A94A31BE02121C CRC64;

Query Match
 Best Local Similarity 100.0%; Score 6; DB 1; Length 219;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 90 AGGLVI 95
 |||||
 DB 31 AGGLVI 36

RESULT 37
 VV_P12H STANDARD: PRT; 225 AA.
 AC P19847;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Nonstructural protein V.
 GN P.V.
 OS Human parainfluenza 2 virus (PIV-2).
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Rubulaviruses.
 OX NCBI_TaxID=11212;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=90281608; PubMed=2162111;
 RX Southern J.A., Precious B., Randall R.E.;
 RT "Two nontemplated nucleotide additions are required to generate the P
 RN mRNA of parainfluenza virus type 2 since the RNA genome encodes

```

RT protein V.;
RL Virology 177:388-390(1990).
CC -I- ALTERNATIVE PRODUCTS: THE P PROTEIN IS COTERMINAL WITH THE V
CC PROTEIN, AND IS ENCODED BY A SEPARATE MRNA WHICH DIFFERS BY TWO
CC NONTEMPLATED NUCLEOTIDES. THE SEQUENCE OF P AND V ARE IDENTICAL
CC UP TO RESIDUE 164.
CC -I- SIMILARITY: TO THE V PROTEIN OF SIMIAN VIRUS 5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M37748; AAA46808.1; -.
DR PIR; A35322; MNNZP2.
DR InterPro: IPR004897; Paramyx_P_V.
DR Pfam; PF03210; Paramyx_P_V; 1.
DR Nonstructural protein; RNA editing.
KM DOMAIN 193 221 CYS-RICH.
FT SEQUENCE 225 AA; 24121 MW; 3F96DA69697F9CCC CRC64;
SQ
Query Match 3.7%; Score 6; DB 1; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 86 PVSAG 91
|11111
Db 148 PVSAG 153
RESULT 38
VV_P12HT STANDARD; PRT; 225 AA.
ID VV_P12HT
AC P23057;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Nonstructural protein V.
GN P/V.
OS Human parainfluenza 2 virus (strain Toshiba) (PIV-2).
OC Viruses: ssRNA negative-strand viruses: Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulaviruses.
OX NCBI_TaxID=11214;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90281574; PubMed=2162103;
RA Ohgimoto S., Bando H., Kawano M., Okamoto K., Tsurudome M.,
RA Nishio M., Ito Y.;
RT "Sequence analysis of P gene of human parainfluenza type 2 virus: P
RT and cysteine-rich proteins are translated by two mRNAs that differ by
RT two nontemplated G residues."
RL Virology 177:116-123(1990).
CC -I- ALTERNATIVE PRODUCTS: THE P PROTEIN IS COTERMINAL WITH THE V
CC PROTEIN, AND IS ENCODED BY A SEPARATE MRNA WHICH DIFFERS BY TWO
CC NONTEMPLATED NUCLEOTIDES. THE SEQUENCE OF P AND V ARE IDENTICAL
CC UP TO RESIDUE 164.
CC -I- SIMILARITY: TO THE V PROTEIN OF SIMIAN VIRUS 5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M37751; AAA46803.1; -.
DR EMBL; X57559; CAA40784.1; -.
DR PIR; A35313; MNNZVT.
DR PIR; S16660; S16660.

```

```

DR InterPro: IPR004897; Paramyx_P_V.
DR Pfam: PF03210; Paramyx_P_V; 1.
DR Nonstructural protein; RNA editing.
KM DOMAIN 193 221 CYS-RICH.
FT SEQUENCE 225 AA; 24151 MW; 3F9D70394CF8DD CRC64;
SQ
Query Match 3.7%; Score 6; DB 1; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 86 PVSAG 91
|11111
Db 148 PVSAG 153
RESULT 39
TRYP_PIG STANDARD; PRT; 231 AA.
ID TRYP_PIG
AC P00761;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin precursor (EC 3.4.21.4).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE OF 1-10.
RA Charles M., Rovey M., Guidoni A.A., Desnuelle P.;
RT "On trypsinogen and trypsin of pig."
RL Biochim. Biophys. Acta 69:115-129(1963).
RN [2]
RP SEQUENCE OF 9-231.
RX MEDLINE=73258692; PubMed=4738933;
RA Hermodson M.A., Ericsson L.H., Neurath H., Walsh K.A.;
RT "Determination of the amino acid sequence of porcine trypsin by
RT sequenator analysis."
RL Biochemistry 12:3146-3153(1973).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX MEDLINE=93187998; PubMed=8445634;
RA Huang Q., Liu S., Tang Y.;
RT "Refined 1.6-A resolution crystal structure of the complex formed
RT between porcine beta-trypsin and MCTI-A, a trypsin inhibitor of the
RT squash family. Detailed comparison with bovine beta-trypsin and its
RT complex."
RL J. Mol. Biol. 229:1022-1030(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX MEDLINE=92201369; PubMed=1551419;
RA Huang Q., Liu S., Tang Y., Zeng F., Qian R.;
RT "Amino acid sequencing of a trypsin inhibitor by refined 1.6 A X-ray
RT crystal structure of its complex with porcine beta-trypsin."
RL FEBS Lett. 297:143-146(1992).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=95035057; PubMed=7947985;
RA Huang Q., Wang Z., Li Y., Liu S., Tang Y.;
RT "Refined 1.8-A resolution crystal structure of the porcine epsilon-
RT trypsin."
RL Biochim. Biophys. Acta 1209:77-82(1994).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF COMPLEX WITH LDTI.
RX MEDLINE=97390427; PubMed=9242660;
RA Stubbs M.T., Morawitz R., Sturzebecher J., Bauer M., Bode W.,
RA Huber R., Piechotka G.P., Matschner G., Sommerhoff C.P., Fritz H.,
RA Auerswald E.A.;
RT "The three-dimensional structure of recombinant leech-derived
RT trypsinase inhibitor in complex with trypsin. Implications for the
RT structure of human mast cell tryptase and its inhibition."
RL J. Biol. Chem. 272:19931-19937(1997).
RN [7]

```

RP X-RAY CRYSTALLOGRAPHY (2.03 ANGSTROMS) OF COMPLEX WITH LDTI.
 RX MEDLINE-98046095; PubMed-9384562;
 RA di Marco S., Priestle J.P.;
 RT "Structure of the complex of leech-derived tryptase inhibitor (LDTI)
 with trypsin and modeling of the LDTI-tryptase system.";
 RL Structure 5:1465-1474(1997).
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-Xaa, Lys-I-Xaa.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 51.
 DR PIR: A00947; TRPTIR.
 DR PDB: 1MCT; 31-JAN-94.
 DR PDB: 1AKS; 12-FEB-97.
 DR PDB: 1EPT; 07-FEB-95.
 DR PDB: 1TFX; 21-JAN-98.
 DR PDB: 1LDT; 20-MAY-98.
 DR PDB: 1ANI; 01-JUL-98.
 DR PDB: 1AVW; 18-NOV-98.
 DR PDB: 1AVX; 18-NOV-98.
 DR MEROPS: S01.151; .
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Ser.protease_Try.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYP-SPEC; 1.
 DR PROSITE: PS50240; TRYPsin_DOM; 1.
 DR PROSITE: PS00134; TRYPsin_HIS; 1.
 DR PROSITE: PS00135; TRYPsin_SER; 1.
 KM Hydrolyase: Serine protease; Digestion; Pancreas; Zymogen;
 KM 3D-structure.
 FT PROPER 1 8
 FT CHAIN 9 231
 FT ACT_SITE 48 48
 FT ACT_SITE 92 92
 FT ACT_SITE 185 185
 FT DISULFID 15 145
 FT DISULFID 33 49
 FT DISULFID 117 218
 FT DISULFID 124 191
 FT DISULFID 156 170
 FT DISULFID 181 205
 FT SITE 179 179
 FT VARIANT 20 20
 FT STRAND 10 10
 FT STRAND 13 14
 FT TURN 17 18
 FT TURN 21 22
 FT STRAND 23 27
 FT STRAND 31 39
 FT TURN 40 41
 FT STRAND 42 45
 FT HELIX 47 49
 FT STRAND 55 58
 FT STRAND 62 62
 FT TURN 63 64
 FT STRAND 71 80
 FT TURN 82 83
 FT TURN 86 88
 FT TURN 90 91
 FT STRAND 94 98
 FT STRAND 112 112
 FT TURN 120 121
 FT STRAND 123 128
 FT STRAND 142 142
 FT STRAND 144 150
 FT HELIX 153 159
 FT TURN 161 163
 FT TURN 166 167
 FT STRAND 168 171
 FT TURN 174 175
 FT STRAND 179 179
 FT TURN 182 183
 FT TURN 185 186
 FT STRAND 188 191

REQUIRED FOR SPECIFICITY (BY SIMILARITY).
 I -> V.

ACTIVATION PEPTIDE.
 TRYPSIN.
 CHARGE RELAY SYSTEM.
 CHARGE RELAY SYSTEM.
 CHARGE RELAY SYSTEM.

FT TURN 192 193
 FT STRAND 194 202
 FT STRAND 207 207
 FT TURN 208 209
 FT STRAND 210 210
 FT STRAND 212 216
 FT HELIX 217 219
 FT HELIX 221 230
 SQ SEQUENCE 231 AA; 24409 MW; A0A125CF7FC138C2 CRC64;
 Query Match 3.7%; Score 6; DB 1; Length 231;
 Best Local Similarity .100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 53 SSSYP 58
 DB 135 SSSYP 140
 RESULT 40
 SODM_MAIZE
 ID SODM_MAIZE STANDARD; PRT; 235 AA.
 AC P09233;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Superoxide dismutase [Mn] 3.1, mitochondrial precursor (EC 1.15.1.1).
 GN SODA.4 OR SOD3.1 OR SOD3.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoidae; Andropogoneae; Zea.
 OX NCBI_TaxId=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. W64A;
 RX MEDLINE-89051020; PubMed-2461225;
 RA Redinbaugh M.G., Madsworth G.T., Scandalios J.G.;
 RT "Isolation and characterization of a cDNA for mitochondrial manganese
 superoxide dismutase (SOD-3) of maize and its relation to other
 RT Biochim. Biophys. Acta 951:61-70(1988).
 RL -1- FUNCTION: Destroys radicals which are normally produced within the
 cells and which are toxic to biological systems.
 CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -1- COFACTOR: Manganese.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X12540; CAA31058.1; .
 DR PIR: S03839; S03839.
 DR HSSP: P04179; IABM.
 DR MAIZEDB: 47587; .
 DR InterPro: IPR001189; SODismutase.
 DR Pfam: PF00081; Sodfe; 1.
 DR Pfam: PF02777; Sodfe_C; 1.
 DR PRODOM: PD000475; SODismutase; 1.
 DR PROSITE: PS00088; SOD_MN; 1.
 KM Oxidoreductase; Manganese; Mitochondrion; Translt peptide;
 KM Multigene family.
 FT TRANSIT 1 31
 FT CHAIN 32 235
 FT METAL 59 59
 MITOCHONDRION.
 SUPEROXIDE DISMUTASE [MN] 3.1.
 MANGANESE (BY SIMILARITY).

FT METAL 107 107 MANGANESE (BY SIMILARITY).
 FT METAL 196 196 MANGANESE (BY SIMILARITY).
 FT METAL 200 200 MANGANESE (BY SIMILARITY).
 SQ SEQUENCE 235 AA; 25545 MW; AD51BAD0F44FDE36 CRC64;

Query Match 3.7%; Score 6; DB 1; Length 235;
 Best Local Similarity 100.0%; Pred. No. 1.le+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 144 TVTLPD 149
 ID 111111
 DB 34 TVTLPD 39

RESULT 41
 CS31_ECOLI STANDARD; PRT; 241 AA.
 ID CS31_ECOLI
 AC P15483;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Chaperone protein CS3-1 precursor.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OX NCBI_TaxID=562;
 RN NCB1
 RP SEQUENCE FROM N.A.
 RC STRAIN=PB176;
 RX MEDLINE=90158116; PubMed=2576094;
 RA Jatakuwari M.B., Thomas C.J., Halter R., Manning P.A.;
 RT "Genes for biosynthesis and assembly of CS3 pili of CFA/II
 RT enterotoxigenic Escherichia coli: novel regulation of pili
 RT production by bypassing an amber codon.";
 RL Mol. Microbiol. 3:1685-1695(1989).
 RN SEQUENCE OF 1-168 FROM N.A.
 RX MEDLINE=89032631; PubMed=2903130;
 RA Boylan M., Smyth C.J., Scott J.R.;
 RT "Nucleotide sequence of the gene encoding the major subunit of CS3
 RT fimbriae of enterotoxigenic Escherichia coli.";
 RL Infect. Immun. 56:3297-3300(1988).
 CC -1- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR THE BIOGENESIS OF
 CC MATURE CS3 PILI.
 CC -1- SUBCELLULAR LOCATION: periplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE PERIPLASMIC PILUS CHAPERONE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X16944; CAA34815.1; -;
 DR EMBL: M35657; -; NOT_ANNOTATED_CDS.
 DR PIR: S07899; S07899.
 DR HSSP: P31697; IPR8.
 DR InterPro: IPR001829; Pili_chaperone.
 DR Pfam: PF00345; Pili_assembly.1.
 DR Pfam: PF02753; Pili_assembly.C.1.
 DR PRINTS: PR00969; CHAPERONPILI.
 DR ProDom: PD001447; Pili_chaperone.1.
 DR PROSITE: PS00635; Pili_CHAPERONE.1.
 DR Chaperone; Fimbria; Periplasmic; signal; Immunoglobulin domain.
 FT SIGNAL 1 21
 FT CHAIN 22 241 CHAPERONE PROTEIN CS3-1.
 FT DISULFID 113 144
 FT CONFLICT 73 73 N -> S (IN REF. 2).
 SQ SEQUENCE 241 AA; 26908 MW; DBC9767D163FCFA0 CRC64;

Query Match 3.7%; Score 6; DB 1; Length 241;
 Best Local Similarity 100.0%; Pred. No. 1.le+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 122 NIYANN 127
 ID 111111
 DB 18 NIYANN 23

RESULT 42
 SDGF_RAT STANDARD; PRT; 243 AA.
 ID SDGF_RAT
 AC P24338;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Schwannoma-derived growth factor precursor (SDGF) (Amphiregulin) (AR).
 GN SDGF.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN NCB1
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=91043095; PubMed=2234093;
 RA Kimura H., Fischer W.H., Schubert D.;
 RT "Structure, expression and function of a schwannoma-derived growth
 RT factor.";
 RL Nature 348:257-260(1990).
 CC -1- FUNCTION: AUTOCRINE GROWTH FACTOR AS WELL AS A MITOGEN FOR A BROAD
 CC RANGE OF TARGET CELLS, AMONG WHICH ASTROCYTES, SCHWANN CELLS AND
 CC FIBROBLASTS.
 CC -1- SIMILARITY: HIGH, TO AMPHIREGULIN.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X55183; CAA38967.1; -;
 DR PIR: S13296; S13296.
 DR HSSP: O99075; IXP1.
 DR InterPro: IPR000561; EGF-like.
 DR Pfam: PF00008; EGF.1.
 DR SMART: SM00181; EGF.1.
 DR PROSITE: PS01186; EGF_2; FALSE_NEG.
 DR PROSITE: PS00022; EGF_1; 1.
 KW Glycoprotein; Cytokine; Growth factor; EGF-like domain; Signal;
 KM Transmembrane.
 FT SIGNAL 1 24
 FT PROPEP 25 96
 FT CHAIN 97 243
 FT DOMAIN 133 173
 FT TRANSMEM 190 213
 FT DISULFID 137 150
 FT DISULFID 145 161
 FT DISULFID 163 172
 FT CARBOHYD 30 30
 FT CARBOHYD 103 103
 FT CARBOHYD 236 236
 SQ SEQUENCE 243 AA; 26633 MW; F59A8173D4921599 CRC64;
 Query Match 3.7%; Score 6; DB 1; Length 243;
 Best Local Similarity 100.0%; Pred. No. 1.le+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 89 SAGGLV 94
 ID 111111
 DB 45 SAGGLV 50

RESULT 43
VNST_PTPV STANDARD: PRT: 250 AA.
ID VNST_PTPV
AC P03516;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nonstructural protein NS-S.
OS Punta toro phlebovirus.
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Phlebovirus.
NC NCBL_TaxID=11587;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84276006; PubMed=6087547;
RA Ihara T., Akashi H., Bishop D.H.L.;
RT "Novel coding strategy (ambisense genomic RNA) revealed by sequence analyses of Punta toro phlebovirus S RNA."
RL Virology 136:293-306(1984).
CC -1- MISCELLANEOUS: THIS PROTEIN MAY BE A TRANSCRIPTASE COMPONENT.
CC -1- SIMILARITY: NS-S FROM PUNTA TORO, RIFT VALLEY FEVER, SANDFLY FEVER SICILIAN, TOSCANA, AND UUKUNIEMI VIRUSES ARE EVOLUTIONARY RELATED.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
DR EMBL: K02736; AAA47115.1; -
DR PIR: A04108; MNVPT.
KM Nonstructural protein; Transcription.
SQ SEQUENCE 250 AA: 29097 MW: 208909A1EDAD90D7 CRC64;

Query Match 3.7%; Score 6; DB 1; Length 250;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 92 GLVIRA 97
Db 80 GLVIRA 85
ID YD37_MYCLE STANDARD: PRT: 251 AA.
AC P53426; Q9C73;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein M1171.
GN M1171 OR B1549_C3_240.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NC NCBL_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX Cole S.T., Elgimeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagsels K., Jancic C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,

RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: STRONG. To M.TUBERCULOSIS RV1337.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U00014; AAA50900.1; -
DR EMBL: AL583921; CAC31552.1; ALT_INIT.
DR Leprosia; M1171; -
DR Interpro: IPR002610; Rhomboid.
DR Pfam: PF01694; Rhomboid; 1.
KM Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 48 68
FT TRANSMEM 88 108
FT TRANSMEM 110 130
FT TRANSMEM 132 152
FT TRANSMEM 158 178
FT TRANSMEM 184 204
FT TRANSMEM 209 229
SQ SEQUENCE 251 AA: 27064 MW: BF847E05059DFC4 CRC64;

Query Match 3.7%; Score 6; DB 1; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 AYGVVL 43
Db 192 AYGVVL 197
ID DSBG_PSEAE STANDARD: PRT: 256 AA.
AC Q91106;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thiol:disulfide interchange protein dsbg precursor.
GN DSBG OR PA2476.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
NC NCBL_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wagman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lartig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -1- FUNCTION: Involved in disulfide bond formation. Functions probably as a disulfide isomerase with a narrower substrate specificity than dsbc. Dsbg is maintained in a reduced state by dsbd (by similarity).
CC -1- SUBUNIT: Homodimer (by similarity).
CC -1- SUBCELLULAR LOCATION: Periplasmic (by similarity).
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY. DSBG SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
CC EMBL; AE004675; AG05864.1; -
CC
CC InterPro: IPR000063; ThioRed.
CC PROSITE; PS00194; THIOREDOXIN; Signal; Complete proteome.
CC Redox-active center; Periplasmic; Signal; Complete proteome.
CC
CC SIGNAL 1 24 POTENTIAL
CC CHAIN 25 256 THIO:DISULFIDE INTERCHANGE PROTEIN DSBG.
CC FT DISULFID 134 137 REDOX-ACTIVE (BY SIMILARITY).
CC SEQUENCE 256 AA; 28053 MW; C1EB26268781CE23 CRC64;
SQ

Query Match 3.7%; Score 6; DB 1; Length 256;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 81 ALYLP 86
| | | | |
Db 65 ALYLP 70

Search completed: November 28, 2002, 19:03:43
Job time : 35 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 18:59:56 ; Search time 81 Seconds
(Without alignments)
409.550 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 161
Sequence: 1 PVTNNGQNLVVDLSTQIFCH.....DVTYLPDYGVPPIPLTVY 161

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	57.1	166	2	Q9F5R9
2	92	57.1	300	2	Q9S497
3	92	57.1	300	2	Q9S496
4	92	57.1	300	2	Q9F550
5	90	55.9	300	16	Q9XBK6
6	73	45.3	300	2	Q9S495
7	59	36.6	300	2	Q9S494
8	59	36.6	300	2	Q9F551
9	59	36.6	300	2	Q9S494
10	56	34.8	300	2	Q9S6R1
11	56	34.8	300	2	Q9F627
12	56	34.8	300	2	Q9A605
13	56	34.8	303	2	Q9A605
14	50	31.1	129	2	Q9XBV8
15	42	26.1	243	2	Q98858
16	9	5.6	295	12	Q998F5

17	9	5.6	304	16	Q8XAX2	Q8XAX2 escherichia
18	9	5.6	550	13	Q9PT94	Q9PT94 brachydanio
19	9	5.6	550	13	Q9DG30	Q9DG30 brachydanio
20	8	5.0	539	16	Q9RMQ4	Q9RMQ4 delnoccocus
21	8	5.0	564	4	Q9H1Y7	Q9H1Y7 homo sapien
22	8	5.0	570	4	Q8WVN2	Q8WVN2 homo sapien
23	8	5.0	599	12	Q80929	Q80929 human papill
24	8	5.0	860	10	Q9SHB7	Q9SHB7 arabidopsis
25	8	5.0	1036	4	Q8WVN1	Q8WVN1 homo sapien
26	8	5.0	1239	5	Q01505	Q01505 caenorhabdi
27	8	5.0	1873	2	Q9Z4N7	Q9Z4N7 enterococcu
28	7	4.3	63	5	Q9HJG6	Q9HJG6 plasmodium
29	7	4.3	73	4	Q75702	Q75702 homo sapien
30	7	4.3	74	5	Q9ND29	Q9ND29 plasmodium
31	7	4.3	87	16	Q31846	Q31846 bacillus su
32	7	4.3	107	5	Q9N3B4	Q9N3B4 caenorhabdi
33	7	4.3	119	10	Q9FE65	Q9FE65 arabidopsis
34	7	4.3	120	10	Q9LJW6	Q9LJW6 arabidopsis
35	7	4.3	128	16	Q9RXH5	Q9RXH5 delnoccocus
36	7	4.3	162	16	Q8UB1	Q8UB1 agrobacteri
37	7	4.3	177	2	Q8VV80	Q8VV80 colwellia m
38	7	4.3	179	5	Q9V7J9	Q9V7J9 plasmodium
39	7	4.3	188	16	Q9PDQ1	Q9PDQ1 xyloella fas
40	7	4.3	210	16	Q9CHD6	Q9CHD6 lactococcus
41	7	4.3	248	16	Q05437	Q05437 mycobacteri
42	7	4.3	248	16	Q8VIR0	Q8VIR0 mycobacteri
43	7	4.3	251	16	Q50195	Q50195 mycobacteri
44	7	4.3	256	16	Q8R805	Q8R805 thermoaer
45	7	4.3	283	10	Q9ZVP6	Q9ZVP6 arabidopsis
46	7	4.3	296	10	Q9LS70	Q9LS70 arabidopsis
47	7	4.3	298	16	Q97IP7	Q97IP7 clostridium
48	7	4.3	304	17	Q9U2U9	Q9U2U9 pyrococcus
49	7	4.3	311	5	Q44789	Q44789 caenorhabdi
50	7	4.3	315	16	Q98141	Q98141 rhizobium l
51	7	4.3	328	10	Q41401	Q41401 sesbania ro
52	7	4.3	365	5	Q9N680	Q9N680 drosophila
53	7	4.3	383	16	Q8ZG16	Q8ZG16 yersinia pe
54	7	4.3	396	2	Q06037	Q06037 lactococcus
55	7	4.3	397	5	Q9V518	Q9V518 drosophila
56	7	4.3	414	16	Q8UGJ7	Q8UGJ7 agrobacteri
57	7	4.3	416	10	Q9LTI6	Q9LTI6 arabidopsis
58	7	4.3	417	2	Q9RPA9	Q9RPA9 streptomyce
59	7	4.3	443	16	Q9CEP2	Q9CEP2 lactococcus
60	7	4.3	457	16	Q9CPE2	Q9CPE2 pasteurella
61	7	4.3	472	10	Q9FRT7	Q9FRT7 arabidopsis
62	7	4.3	478	11	Q9D306	Q9D306 mus musculu
63	7	4.3	478	11	Q9D2X2	Q9D2X2 mus musculu
64	7	4.3	500	17	Q978P9	Q978P9 thermoplas
65	7	4.3	523	2	Q9KHJ9	Q9KHJ9 streptomyce
66	7	4.3	524	16	Q942P2	Q942P2 oryza sativ
67	7	4.3	533	10	Q8U6S7	Q8U6S7 agrobacteri
68	7	4.3	569	10	Q9FV79	Q9FV79 arabidopsis
69	7	4.3	590	16	Q9RWT8	Q9RWT8 delnoccocus
70	7	4.3	604	5	Q8SVL2	Q8SVL2 drosophila
71	7	4.3	629	13	Q90WR5	Q90WR5 lampetra fl
72	7	4.3	650	17	Q9HP27	Q9HP27 halobacteri
73	7	4.3	656	17	Q97A14	Q97A14 thermoplas
74	7	4.3	687	16	P72758	P72758 synchocyst
75	7	4.3	696	5	Q18344	Q18344 caenorhabdi
76	7	4.3	707	2	Q9X5B5	Q9X5B5 neisseria m
77	7	4.3	714	16	Q8XV60	Q8XV60 ralatonia s
78	7	4.3	722	2	Q9X5B4	Q9X5B4 neisseria m
79	7	4.3	741	16	Q9JTK3	Q9JTK3 neisseria m
80	7	4.3	761	16	Q8X893	Q8X893 escherichia
81	7	4.3	839	16	Q97KK0	Q97KK0 clostridium
82	7	4.3	866	16	Q9LJ33	Q9LJ33 streptomyce
83	7	4.3	1001	10	Q9MAT0	Q9MAT0 arabidopsis
84	7	4.3	1004	10	Q48844	Q48844 arabidopsis
85	7	4.3	1040	16	Q8ZM02	Q8ZM02 salmone
86	7	4.3	1040	16	Q8ZSE7	Q8ZSE7 salmone
87	7	4.3	1040	16	Q8XJ34	Q8XJ34 escherichia
88	7	4.3	1041	16	Q8XQ04	Q8XQ04 ralatonia s
89	7	4.3	1043	16	Q910V6	Q910V6 pseudomonas


```

966 6 3.7 693 3 09C4A2
967 6 3.7 693 12 069090
968 6 3.7 695 5 09MAC4
969 6 3.7 695 16 08XR06
970 6 3.7 696 10 09AD26
971 6 3.7 696 17 08ZTR8
972 6 3.7 698 17 026996
973 6 3.7 700 3 08TG00
974 6 3.7 702 3 074312
975 6 3.7 710 12 08V4W8
976 6 3.7 710 12 08V2P8
977 6 3.7 710 12 08QMU3
978 6 3.7 711 12 09DHU5
979 6 3.7 711 12 08V3K1
980 6 3.7 714 12 091MR5
981 6 3.7 715 10 08W1S4
982 6 3.7 716 13 091691
983 6 3.7 717 13 091691
984 6 3.7 718 16 09P0K8
985 6 3.7 719 16 09AB85
986 6 3.7 720 3 014367
987 6 3.7 720 5 09GRN6
988 6 3.7 725 2 052978
989 6 3.7 725 5 09UIE0
990 6 3.7 726 2 09AUB3
991 6 3.7 727 2 019327
992 6 3.7 727 10 09T0B4
993 6 3.7 727 10 08SA09
994 6 3.7 730 2 095NF2
995 6 3.7 730 10 065736
996 6 3.7 732 17 0970T3
997 6 3.7 734 16 09HT76
998 6 3.7 737 10 09M2Y4
999 6 3.7 737 10 08S6T7
1000 6 3.7 737 10 08S6T7

```

ALIGNMENTS

```

RESULT 1
ID 09F5R9 PRELIMINARY: PRT: 166 AA.
AC 09F5R9:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE FIMH.
GN FIMH.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12:
RA Harris S.L., Spears P.A., Havelle E.A., Hamrick T.S., Horton J.R.,
RT "Isolation and characterization of Escherichia coli type 1 pili
mutants that have altered binding specificities."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF306537; AAG30927.1; -.
SQ SEQUENCE 166 AA; 17826 MW; 2E64EEA80AE286DC CRC64;

```

```

Query Match 57.1%; Score 92; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.2e-88;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 PNVVNGONLVVDSTOIFCHNDYPERITDVTYTLORGSAVGVLSNFGVYKSGSSYPP 60
DB 47 PNVVNGONLVVDSTOIFCHNDYPERITDVTYTLORGSAVGVLSNFGVYKSGSSYPP 106
QY 61 TTSETPRVYNSRTDKPMPVALYLTTPVSSAG 92

```

```

DB 107 TTSETPRVYNSRTDKPMPVALYLTTPVSSAG 138

```

RESULT 2

```

ID 09S497 PRELIMINARY: PRT: 300 AA.
AC 09S497:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE FIMH.
GN FIMH.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12:
RA Hamrick T.S., Harris S.L., Spears P.A., Havelle E.A., Horton J.R.,
RT "Genetic characterization of Escherichia coli type 1 pilus adhesin
mutants and identification of a novel binding phenotype."
RL J. Bacteriol. 182:4012-4021(2000).
DR EMBL: AF154925; AAD44319.1; -.
DR InterPro: IPR000259; Fimbrin.
DR Pfam: PF00419; Fimbrin; 1.
FT VARIANT 19 19 D -> A.
SQ SEQUENCE 300 AA; 31517 MW; D29340A1575A2358 CRC64;

```

```

Query Match 57.1%; Score 92; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 3.7e-88;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 PNVVNGONLVVDSTOIFCHNDYPERITDVTYTLORGSAVGVLSNFGVYKSGSSYPP 60
DB 47 PNVVNGONLVVDSTOIFCHNDYPERITDVTYTLORGSAVGVLSNFGVYKSGSSYPP 106
QY 61 TTSETPRVYNSRTDKPMPVALYLTTPVSSAG 92
DB 107 TTSETPRVYNSRTDKPMPVALYLTTPVSSAG 138

```

RESULT 3

```

ID 09S496 PRELIMINARY: PRT: 300 AA.
AC 09S496:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE FIMH.
GN FIMH.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12:
RA Hamrick T.S., Harris S.L., Spears P.A., Havelle E.A., Horton J.R.,
RT "Genetic characterization of Escherichia coli type 1 pilus adhesin
mutants and identification of a novel binding phenotype."
RL J. Bacteriol. 182:4012-4021(2000).
DR EMBL: AF154926; AAD44320.1; -.
DR InterPro: IPR000259; Fimbrin.
DR Pfam: PF00419; Fimbrin; 1.
FT VARIANT 21 21 L -> S.
SQ SEQUENCE 300 AA; 31499 MW; 3ADP97A85F8CE478 CRC64;

```


Query Match 57.1%; Score 92; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 3.7e-88;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVAWGNLVVDSTQIFCHNDYPTTIDYVTLQSGAYGVLNFGSVTKYSGSSYPFP 60
DB 47 PVAWGNLVVDSTQIFCHNDYPTTIDYVTLQSGAYGVLNFGSVTKYSGSSYPFP 106

OY 61 TTSETPRVYVNSRTDKPMPVALYLTTPVSSAG 92
DB 107 TTSETPRVYVNSRTDKPMPVALYLTTPVSSAG 138

RESULT 4
O9F5S0 PRELIMINARY; PRT: 300 AA.
ID O9F5S0;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
GN FIMH.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA MEDLINE-21289098; PubMed-11395476;
RA Harris S.L., Spears P.A., Havelle E.A., Hamrick T.S., Horton J.R.,
RT "Characterization of Escherichia coli Type 1 Pilus Mutants with
Altered Binding Specificities";
RL J. Bacteriol. 183:4099-4102(2001).
DR EMBL: AF306536; AAG30926.1; -;
DR InterPro: IPR000259; Fimbril.
DR Pfam: PF00419; Fimbril; 1.
FT VARIANT 163 163 S -> P.
SQ SEQUENCE 300 AA; 31413 MW; 922904051758746D CRC64;

Query Match 57.1%; Score 92; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 3.7e-88;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVAWGNLVVDSTQIFCHNDYPTTIDYVTLQSGAYGVLNFGSVTKYSGSSYPFP 60
DB 47 PVAWGNLVVDSTQIFCHNDYPTTIDYVTLQSGAYGVLNFGSVTKYSGSSYPFP 106

OY 61 TTSETPRVYVNSRTDKPMPVALYLTTPVSSAG 92
DB 107 TTSETPRVYVNSRTDKPMPVALYLTTPVSSAG 138

RESULT 5
O8XBA6 PRELIMINARY; PRT: 300 AA.
ID O8XBA6;
AC O8XBA6;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
GN Minor fimbrial subunit, D-mannose specific adhesin.
DE FIMH OR Z5918 OR ECS5279.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RA MEDLINE-21074935; PubMed-11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Posfai G., Hackett J., Klink S., Routin A., Shao Y., Miller L.,
RA Grobbeck E.J., Davis N.W., Lim A., Dinalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.,
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7,"
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE-21156231; PubMed-11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Ikeda T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.,
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12,"
RL DNA Res. 8:11-22(2001).
DR EMBL: AF005662; AAG59502.1; -;
DR EMBL: AF002569; BAB38702.1; -;
DR InterPro: IPR000259; Fimbril.
DR Pfam: PF00419; Fimbril; 1.
SQ COMPLETE proteome.
SQ SEQUENCE 300 AA; 31459 MW; 68FE713A3D372CB CRC64;

Query Match 55.9%; Score 90; DB 16; Length 300;
Best Local Similarity 100.0%; Pred. No. 4.7e-86;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VNWGNLVVDSTQIFCHNDYPTTIDYVTLQSGAYGVLNFGSVTKYSGSSYPFP 62
DB 49 VNWGNLVVDSTQIFCHNDYPTTIDYVTLQSGAYGVLNFGSVTKYSGSSYPFP 108

OY 63 SETPRVYVNSRTDKPMPVALYLTTPVSSAG 92
DB 109 SETPRVYVNSRTDKPMPVALYLTTPVSSAG 138

RESULT 6
O9S495 PRELIMINARY; PRT: 300 AA.
ID O9S495;
AC O9S495;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
GN FIMH.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-20327582; PubMed-10869080;
RA Hamrick T.S., Harris S.L., Spears P.A., Havelle E.A., Horton J.R.,
RA Russell P.W., Orndorff P.E.,
RT "Genetic characterization of Escherichia coli type 1 pilus adhesin
mutants and identification of a novel binding phenotype,"
RL J. Bacteriol. 182:4012-4021(2000).
DR EMBL: AF154927; AAD44321.1; -;
DR InterPro: IPR000259; Fimbril.
DR Pfam: PF00419; Fimbril; 1.
FT VARIANT 65 65 Y -> C.
SQ SEQUENCE 300 AA; 31533 MW; 48FEB2795A3CB23D CRC64;

Query Match 45.3%; Score 73; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 3.4e-68;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 HNDYPTTIDYVTLQSGAYGVLNFGSVTKYSGSSYPFP 79
DB 66 HNDYPTTIDYVTLQSGAYGVLNFGSVTKYSGSSYPFP 125

OY 80 VALYLPVSSAG 92
|
Db 126 VALYLPVSSAG 138

RESULT 7

09S494 ID 09S494 PRELIMINARY; PRT; 300 AA.
AC 09S494;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE FIMH.
GN Escherichia coli.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=20327582; PubMed=10869080;
RA Hamrick T.S., Harris S.L., Spears P.A., Havelle E.A., Horton J.R.,
RA Russell P.W., Orndorff P.E.;
RT "Genetic characterization of Escherichia coli type 1 pilus adhesin
mutants and identification of a novel binding phenotype.";
RL J. Bacteriol. 182:4012-4021(2000).
DR EMBL: AF154928; AAD4322.1;
DR InterPro: IPR000259; Fimbril.
DR Pfam: PF00419; Fimbril; 1.
FT VARIANT 79 79 R -> L.
SQ SEQUENCE 300 AA; 31516 MW; B6073DF68366B86E CRC64;

Query Match 36.6%; Score 59; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.7e-53;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 ORGSAYGVLSNFGTGYKSSGSSYPPTTSETPRVVYNSRTDKPMPVALYLPVSSAG 92
|
Db 80 ORGSAYGVLSNFGTGYKSSGSSYPPTTSETPRVVYNSRTDKPMPVALYLPVSSAG 138

RESULT 8
09F5S1 ID 09F5S1 PRELIMINARY; PRT; 300 AA.

AC 09F5S1;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE FIMH.
GN Escherichia coli.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=21289098; PubMed=11395476;
RA Harris S.L., Spears P.A., Havelle E.A., Hamrick T.S., Horton J.R.,
RA Orndorff P.E.;
RT "Characterization of Escherichia coli Type 1 Pilus Mutants with
Altered Binding Specificities";
RL J. Bacteriol. 183:4099-4102(2001).
DR EMBL: AF306535; AAG30925.1;
DR InterPro: IPR000259; Fimbril.
DR Pfam: PF00419; Fimbril; 1.
FT VARIANT 79 79 P -> R.
SQ SEQUENCE 300 AA; 31457 MW; 9F7B0A35FF51F938 CRC64;

Query Match 36.6%; Score 59; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.7e-53;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 ORGSAYGVLSNFGTGYKSSGSSYPPTTSETPRVVYNSRTDKPMPVALYLPVSSAG 92
|
Db 80 ORGSAYGVLSNFGTGYKSSGSSYPPTTSETPRVVYNSRTDKPMPVALYLPVSSAG 138

RESULT 9

09R5Y2 ID 09R5Y2 PRELIMINARY; PRT; 300 AA.
AC 09R5Y2;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE FIMH SUBUNIT-MANNOSE-sensitive type 1 fimbrial adhesin.
GN Escherichia coli.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94131954; PubMed=7905476;
RX Sokurenko E.V., Courtney H.S., Ohman D.E., Klemm P., Hasty D.L.;
RA J. Bacteriol. 176:748-755(1994).
DR InterPro: IPR000259; Fimbril.
DR Pfam: PF00419; Fimbril; 1.
SQ SEQUENCE 300 AA; 31552 MW; E75328B7C366A319 CRC64;

Query Match 36.6%; Score 59; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.7e-53;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 ORGSAYGVLSNFGTGYKSSGSSYPPTTSETPRVVYNSRTDKPMPVALYLPVSSAG 92
|
Db 80 ORGSAYGVLSNFGTGYKSSGSSYPPTTSETPRVVYNSRTDKPMPVALYLPVSSAG 138

RESULT 10

09S6R1 ID 09S6R1 PRELIMINARY; PRT; 300 AA.
AC 09S6R1;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Type 1 fimbriae adhesin, precursor polypeptide.
GN Escherichia coli.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT78;
RX MEDLINE=88038337; PubMed=2890081;
RA Klemm P., Christiansen G.;
RT "Three fim genes required for the regulation of length and mediation
of adhesion of Escherichia coli type 1 fimbriae";
RL Mol. Gen. Genet. 208:439-445(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MT78;
RX MEDLINE=96235989; PubMed=8636962;
RA Marc D., Dho-Moulin M.;
RT "Analysis of the fim cluster of an avian O2 strain of Escherichia
coli: serogroup-specific sites within fima and nucleotide sequence of
fimI";
RL J. Med. Microbiol. 44:444-452(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-MT78;
RX MEDLINE=96125229; PubMed=8550444;
RA Klemm P., Tong S., Nielsen H., Conway T.;
RT "The gfp gene of Escherichia coli involved in gluconate uptake";
RL J. Bacteriol. 178:61-67(1996).

[4]
RN SEQUENCE FROM N.A.
RP STRAIN-MT78;
RC MEDLINE-97426617; PubMed-9278503;
RA Blatter F.R., Plunkett G.J.II, Bloch C.A., Perna N.T., Burland V.,
Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474(1997).
[5]
RN SEQUENCE FROM N.A.
RP STRAIN-MT78;
RC MEDLINE-98439336; PubMed-9766199;
RA Marc D., Arne P., Bree A., Dho-Moulin M.;
RT "Colonization ability and pathogenic properties of a fim- mutant of an
avian strain of *Escherichia coli*.";
RL Res. Microbiol. 149:473-485(1998).
DR EMBL: AJ225176; CAI2423.1;
DR InterPro: IPR000259; Fimbril.
DR Pfam: PF00419; Fimbril; 1.
SQ SEQUENCE 300 AA; 31459 MW; EA040CEDD2137FE8 CRC64;

Query Match 34.8%; Score 56; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.4e-50;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 IKAGSLIAVLILRQTNNYNSDDQFVWNIYANDVVPYPTGGCDVSARDVTVTLDPY 150
Db 141 IKAGSLIAVLILRQTNNYNSDDQFVWNIYANDVVPYPTGGCDVSARDVTVTLDPY 196

RESULT 11
087634
ID 087634 PRELIMINARY; PRT; 300 AA.
AC 087634;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE FimH.
GN FimH.
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IHE3034;
RA Poutu R., Korhonen T.K.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF089840; AAC35864.1;
DR InterPro: IPR000259; Fimbril.
DR Pfam: PF00419; Fimbril; 1.
SQ SEQUENCE 300 AA; 31429 MW; D789457A5F5255AE CRC64;

Query Match 34.8%; Score 56; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.4e-50;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 IKAGSLIAVLILRQTNNYNSDDQFVWNIYANDVVPYPTGGCDVSARDVTVTLDPY 150
Db 141 IKAGSLIAVLILRQTNNYNSDDQFVWNIYANDVVPYPTGGCDVSARDVTVTLDPY 196

RESULT 12
09F627
ID 09F627 PRELIMINARY; PRT; 300 AA.
AC 09F627;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Type 1 fimbrial adhesin subunit FimH.
GN FimH.
OX

OS *Escherichia coli*.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LF82;
RA MEDLINE-21150472; PubMed-11251843;
RT Boudreau J., Barnich N., Dartevelle-Michaud A.;
RT "Type 1 pili-mediated adherence of *Escherichia coli* strain LF82
isolated from Crohn's disease is involved in bacterial invasion of
intestinal epithelial cells.";
RL Mol. Microbiol. 39:1272-1284(2001).
DR EMBL: AF288194; AAC24827.1;
DR InterPro: IPR000259; Fimbril.
DR Pfam: PF00419; Fimbril; 1.
SQ SEQUENCE 300 AA; 31445 MW; BA040E9DD2137E43 CRC64;

Query Match 34.8%; Score 56; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.4e-50;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 IKAGSLIAVLILRQTNNYNSDDQFVWNIYANDVVPYPTGGCDVSARDVTVTLDPY 150
Db 141 IKAGSLIAVLILRQTNNYNSDDQFVWNIYANDVVPYPTGGCDVSARDVTVTLDPY 196

RESULT 13
09AP05
ID 09AP05 PRELIMINARY; PRT; 303 AA.
AC 09AP05;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE FimH.
GN FimH.
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RS218;
RA Elliott S.J., Kim K.S.;
RL "Virulence factors of meningitic *Escherichia coli*.";
RT Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF317710; AAG60686.1;
DR InterPro: IPR000259; Fimbril.
DR Pfam: PF00419; Fimbril; 1.
SQ SEQUENCE 303 AA; 31773 MW; 4F615EASD5E6763C CRC64;

Query Match 34.8%; Score 56; DB 2; Length 303;
Best Local Similarity 100.0%; Pred. No. 2.5e-50;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 IKAGSLIAVLILRQTNNYNSDDQFVWNIYANDVVPYPTGGCDVSARDVTVTLDPY 150
Db 144 IKAGSLIAVLILRQTNNYNSDDQFVWNIYANDVVPYPTGGCDVSARDVTVTLDPY 199

RESULT 14
09XBV8
ID 09XBV8 PRELIMINARY; PRT; 129 AA.
AC 09XBV8;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE FimH.
GN *Escherichia coli*.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;

[1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA MEDLINE-20327582; PubMed=10869080;
 RA Hamrick T.S., Harris S.L., Spears P.A., Havelle E.A., Horton J.R.,
 RA Russell P.W., Orendorff P.E.;
 RA "Genetic characterization of *Escherichia coli* type 1 pilus adhesin
 RT mutants and identification of a novel binding phenolype-";
 RL J. Bacteriol. 182:4012-4021(2000).
 DR EMBL: AF154929; AAD44323.1; -;
 SO SEQUENCE 129 AA; 14142 MW; BDEID4D0E1EC536 CRC64;

Query Match 31.1%; Score 50; DB 2; Length 129;
 Best Local Similarity 100.0%; Pred. No. 2.3e-44;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVVNGQNLVVDLSTQIFCHNDYPTITDYVTLQSGAVGVLNFGTV 50
 ||||||||||||||||||||||||||||||||||||||||||||
 DB 47 PVVNGQNLVVDLSTQIFCHNDYPTITDYVTLQSGAVGVLNFGTV 96

RESULT 15
 Q08858
 ID Q08858 PRELIMINARY; PRT; 243 AA.
 AC Q08858;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 06, Last annotation update)
 DE FIMH protein precursor.
 GN FIMH.
 OS *Klebsiella pneumoniae*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Klebsiella*.
 OX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-89155420; PubMed=2563996;
 RA Gerlach G.F., Clegg S., Allen B.L.;
 RT "Identification and characterization of the genes encoding the type 3
 RT and type 1 fimbrial adhesins of *Klebsiella pneumoniae*.";
 RL J. Bacteriol. 171:1262-1270(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Nichols N.N., Clegg S.;
 RL Submitted (Aug-1993) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN REGULATION OF LENGTH AND MEDIATION OF
 CC ADHESION OF TYPE 1 FIMBRIAE (BUT NOT NECESSARY FOR THE PRODUCTION
 CC OF FIMBRIAE).
 CC EMBL: L23111; AAA25063.1; -;
 DR FIMBria; Signal.
 KW FIMBria; Signal.
 FT SIGNAL 25
 FT CHAIN 1 243 FIMH PROTEIN.
 SO SEQUENCE 243 AA; 25977 MW; C8394DVC44A98AE7 CRC64;

Query Match 26.1%; Score 42; DB 2; Length 243;
 Best Local Similarity 100.0%; Pred. No. 1e-35;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VNVGQNLVVDLSTQIFCHNDYPTITDYVTLQSGAVGVLN 44
 ||||||||||||||||||||||||||||||||||||||||||||
 DB 51 VNVGQNLVVDLSTQIFCHNDYPTITDYVTLQSGAVGVLN 92

RESULT 16
 Q098F5
 ID Q098F5 PRELIMINARY; PRT; 295 AA.
 AC Q098F5;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)
 DE Probable capsid protein.
 OS porcine cytomegalovirus.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Betaherpesvirinae; Cytomegalovirus.
 OX NCBI_TaxID=109993;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OF-1;
 RA Rupasingha V., Iwatsuki-Horimoto K., Tajima T., Sugii S., Horimoto T.;
 RT "Analysis of the porcine cytomegalovirus major capsid protein gene.";
 RL Submitted (Nov-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB051069; BAB40661.1; -;
 DR InterPro: IPR002690; Herpes_V23.
 DR Pfam: PF01802; Herpes_V23; 1.
 SO SEQUENCE 295 AA; 33392 MW; 953641F47D07E620 CRC64;

Query Match 5.6%; Score 9; DB 12; Length 295;
 Best Local Similarity 100.0%; Pred. No. 0.57;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 141 RDVYVTLPD 149
 ||||||||
 DB 172 RDVYVTLPD 180

RESULT 17
 Q08XAX2
 ID Q08XAX2 PRELIMINARY; PRT; 304 AA.
 AC Q08XAX2;
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Putative adhesin, similar to FIMH protein.
 GN 22206 OR ECS2107.
 OS *Escherichia coli* O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Escherichia*.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RC MEDLINE-21074935; PubMed=1120651;
 RX Pena N.T., Plunkett G., Iii, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Iam A., Dimailanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner P.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE-21156231; PubMed=1128796;
 RA Hayashi T., Makino K., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohnishi M., Murata T., Tanaka M., Toke T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 DR EMBL: AE005354; AAC56266.1; -;
 DR EMBL: AP002557; BAB35530.1; -;
 DR InterPro: IPR000259; Fimbrin1.
 DR Pfam: PF00419; Fimbrin1; 1.
 KW Complete proteome.
 SO SEQUENCE 304 AA; 32041 MW; EED538023D95AFD5 CRC64;

Query Match 5.6%; Score 9; DB 16; Length 304;
 Best Local Similarity 100.0%; Pred. No. 0.59;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 GQNLVVDLS 14
 ||||||||
 DB 53 GQNLVVDLS 61

```

RESULT 18
Q9PT94      PRELIMINARY;      PRT;      550 AA.
ID  Q9PT94
AC  Q9PT94
DT  01-MAY-2000 (TRENBLREL. 13, Created)
DT  01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DE  01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE  Winged helix nude.
GN  WHN.
OS  Brachydanio rerio (zebrafish) (zebra danio).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC  Cyprinidae; Danio.
OX  NCBI_Taxid=7955;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE-20509988; PubMed-11054532;
RA  Schlake T., Schorpp M., Boehm T.;
RT  "Formation of regulator/target gene relationships during evolution.";
RL  Gene 236:29-34(2000).
DR  EMBL; AJ252024; CAB64948.1; -.
DR  HSSP; Q63245; 2HFH.
DR  InterPro; IPR001766; TF_Fork_head.
DR  Pfam; PF00250; FORKHEAD.
DR  PRINTS; PR00053; FORKHEAD.
DR  ProDom; PD000425; TF_Fork_head.1.
DR  SMART; SM00339; FH; 1.
DR  PROSITE; PS00658; FORK_HEAD_2; 1.
DR  PROSITE; PS50039; FORK_HEAD_3; 1.
SQ  SEQUENCE 550 AA; 60963 MW; A23748F2022B3792 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 13; Length 550;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  85 TPVSSAGCL 93
DB  505 TPVSSAGCL 513

RESULT 19
Q9DG30      PRELIMINARY;      PRT;      550 AA.
ID  Q9DG30
AC  Q9DG30
DT  01-MAR-2001 (TRENBLREL. 16, Created)
DT  01-DEC-2001 (TRENBLREL. 19, Last sequence update)
DT  01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE  Transcription factor foxn4.
GN  FOXN4.
OS  Brachydanio rerio (zebrafish) (zebra danio).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC  Cyprinidae; Danio.
OX  NCBI_Taxid=7955;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  STRAIN-TUEBINGEN; TISSUE-THYMUS;
RA  Vissel A., Willlett C.E., Danilova N.P., Steiner L.A.;
RT  "Expression of winged-helix transcription factor foxn4 (whn) in
RT  zebrafish."
RL  Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF198446; AAC27086.2; -.
DR  HSSP; Q63245; 2HFH.
DR  InterPro; IPR001766; TF_Fork_head.
DR  Pfam; PF00250; Fork_head.2.
DR  PRINTS; PR00053; FORKHEAD.
DR  ProDom; PD000425; TF_Fork_head.1.
DR  SMART; SM00339; FH; 2.
DR  PROSITE; PS00658; FORK_HEAD_2; 2.
DR  PROSITE; PS50039; FORK_HEAD_3; 2.
SQ  SEQUENCE 550 AA; 60977 MW; A8E3D8F2022B3797 CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Score 9; DB 13; Length 550;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  85 TPVSSAGCL 93
DB  505 TPVSSAGCL 513

RESULT 20
Q9RMQ4      PRELIMINARY;      PRT;      539 AA.
ID  Q9RMQ4
AC  Q9RMQ4
DT  01-MAY-2000 (TRENBLREL. 13, Created)
DT  01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT  01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE  Arginine utilization protein ROCB, putative.
GN  DR0612.
OS  Deinococcus radiodurans.
OC  Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC  Deinococcaceae; Deinococcus.
OX  NCBI_Taxid=1299;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE-20036896; PubMed-10567266;
RA  White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA  Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
RA  Moffatt K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA  Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA  Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA  Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA  Fraser C.M.;
RT  "Genome sequence of the radioresistant bacterium Deinococcus
RT  radiodurans R1."
RL  Science 286:1571-1577(1999).
DR  EMBL; AE001919; AMF10189.1; -.
DR  TIGR; DR0612; -.
DR  InterPro; IPR002933; Peptidase_M20.
DR  Pfam; PF01546; Peptidase_M20; 1.
KW  Complete proteome.
SQ  SEQUENCE 539 AA; 57561 MW; 946BFD6F7EDF37B CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 539;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  137 DVSARDVT 144
DB  4 DVSARDVT 11

RESULT 21
Q9H1Y7      PRELIMINARY;      PRT;      564 AA.
ID  Q9H1Y7
AC  Q9H1Y7
DT  01-MAR-2001 (TRENBLREL. 16, Created)
DT  01-MAR-2001 (TRENBLREL. 16, Last sequence update)
DT  01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE  D1862P8.3 (Similar to MAP3K10 (Mitogen-activated protein kinase kinase
DE  kinase 10)) (Fragment).
GN  D1862P8.3.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX  NCBI_Taxid=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Hall R.;
RA  Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC  -1 SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR  EMBL; AL133380; CAC17571.1; -.
DR  HSSP; P06241; 1SHF.

```

DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR ProDom: PD000066; SH3; 1.
 DR SMART: SM00326; SH3; 1.
 DR SMART: SM00220; S_TKC; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS50002; SH3; 1.
 DR Kinase: SH3 domain.
 KW NON_TER
 FT SEQUENCE 564 AA; 62400 MW; DAF91E14842E7EEB CRC64;

Query Match 5.0%; Score 8; DB 4; Length 564;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 TPVSSAG 92
 |||||
 Db 12 TPVSSAG 19

RESULT 22

O8WMN2 PRELIMINARY; PRT; 570 AA.

AC O8WMN2:
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Mixed lineage kinase Aalpa.
 GN MLKALPHA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HEART;
 RA Krashu S., Protopopov A., Rynditch A., Zabarovsky E., Kashuba V.;
 RT "MLK4, a new member of mixed lineage kinases."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ311797; CAC84639.1;
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR ProDom: PD000066; SH3; 1.
 DR SMART: SM00326; SH3; 1.
 DR SMART: SM00220; S_TKC; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN.1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; UNKNOWN.1.
 DR PROSITE: PS50002; SH3; 1.
 KW kinase
 FT SEQUENCE 570 AA; 62922 MW; 34CE937FCBCC5F93 CRC64;

Query Match 5.0%; Score 8; DB 4; Length 570;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 TPVSSAG 92
 |||||
 Db 12 TPVSSAG 19

RESULT 23

O80929 PRELIMINARY; PRT; 599 AA.

AC O80929:
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Replication protein El.
 GN El.
 OS Human papillomavirus type 50.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_Taxid=40539;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Delius H.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Farmer A.D.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF
 CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2
 CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH
 CC CONTAINS BINDING SITES FOR BOTH PROTEINS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 DR EMBL: U31790; AAA79473.1;
 DR InterPro: IPR001177; Papillom_El.
 DR Pfam: PF00519; El; 1.
 DR Pfam: PF00524; ELN; 1.
 DR Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;
 KW Nuclear protein.
 FT NP_BIND 427 434 ATP (POTENTIAL).
 FT SEQUENCE 599 AA; 68862 MW; EABD2ED3E8C63099 CRC64;

Query Match 5.0%; Score 8; DB 12; Length 599;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 109 TNNNSDD 116
 |||||
 Db 381 TNNNSDD 388

RESULT 24

O9SHB7 PRELIMINARY; PRT; 860 AA.

AC O9SHB7:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE At2g40720 protein.
 GN AT2G40720.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., VanAken S.E., Unayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,

RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:761-768(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RT Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC007660; AAD32807.1; -
 DR InterPro: IPR000873; AMP-bind.
 DR InterPro: IPR002885; PPR.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF01535; PPR; 18.
 DR TIGRFAMs: TIGR00756; PPR; 12.
 DR PROSITE: PS00455; AMP_BINDING; UNKNOWN_1.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
 SQ SEQUENCE 860 AA; 95393 MW; A79159C501D77976 CRC64;

Query Match 5.0%; Score 8; DB 10; Length 860;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 138 VSARDVTY 145
 Db 128 VSARDVTY 135

RESULT 25
 08WMN1 PRELIMINARY; PRT; 1036 AA.
 ID 08WMN1
 AC 08WMN1
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Mixed lineage kinase Abeta.
 GN MK4BETA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Krasna S., Protopopov A., Rynditch A., Zabarovsky E., Kashuba V.;
 RT "MLK, a new member of mixed lineage kinases.";
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ311798; CAC84640.1; -
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR001452; SH3
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00326; SH3; 1.
 DR SMART: SM00220; S_TKc; 1.
 DR SMART: SM00219; TYRc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 DR PROSITE: PS00108; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00002; SH3; 1.
 KW kinase.
 SQ SEQUENCE 1036 AA; 113809 MW; B9C2ACF397C2CF37 CRC64;

Query Match 5.0%; Score 8; DB 4; Length 1036;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 85 TPVSSAGG 92
 Db 11111111

Db 12 TPVSSAGG 19

RESULT 26
 001505 PRELIMINARY; PRT; 1239 AA.
 ID 001505
 AC 001505
 DT 01-JUL-1997 (TREMBlrel. 04, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Hypothetical 135.4 kDa protein.
 GN C37A2.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Le T.T., Kemp K., Scheet P.;
 RT "The sequence of C. elegans cosmid C37A2.";
 RL Submitted (APR-1997) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RT "Direct Submission.";
 RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: U97194; AAB52447.2; -
 DR InterPro: IPR002965; P-rich_extensn.
 DR PRINTS: PR01217; PRICHEXTENS.
 KW Hypothetical protein.
 SQ SEQUENCE 1239 AA; 135436 MW; 81F6C299FDC77D85 CRC64;

Query Match 5.0%; Score 8; DB 5; Length 1239;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 151 RGSVPDPL 158
 Db 960 RGSVPDPL 967

RESULT 27
 0924N7 PRELIMINARY; PRT; 1873 AA.
 ID 0924N7
 AC 0924N7
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Surface protein precursor.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Enterococcaceae; Enterococcus.
 ON NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MH594;
 RX MEDLINE=99081742; PubMed=9864215;
 RA Shankar V., Baghdayan A.S., Huycke M.M., Lindahl G., Gilmore M.S.;
 RT "Infection-derived Enterococcus faecalis strains are enriched in esp,
 RL a gene encoding a novel surface protein.";
 DR EMBL: AF034779; AAD09858.1; -
 DR InterPro: IPR001899; Gram_pos_anchor.

DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR TIGRFRAMS: TIGR01167; LPXTG_anchor; 1.
 DR TIGRFRAMS: TIGR01168; XSRK_signal; 1.
 KW Signal.
 FT SIGNAL 1 49 POTENTIAL.
 FT CHAIN 50 1873 SURFACE PROTEIN.
 SQ SEQUENCE 1873 AA; 202084 MW; F609483DB9B80621 CRC64;

Query Match 5.0%; Score 8; DB 2; Length 1873;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 DVTVLPD 149
 |||||
 DB 1754 DVTVLPD 1761

RESULT 28
 O9BJG6 PRELIMINARY; PRT; 63 AA.
 AC O9BJG6:
 DT 01-JUN-2001 (TREMBlrel. 17; Created)
 DT 01-JUN-2001 (TREMBlrel. 17; Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17; Last annotation update)
 DE Variant surface protein (Fragment).
 GN VAR.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=209;
 RA Nogueira P.A., Wunderlich G., Tada M.S., Costa J.D.N., Menezes M.J.,
 RA Scherf A., Pereira da Silva L.H.;
 RT "Plasmodium falciparum: repertoire of expressed var genes and adhesion
 RT properties to endothelial receptors of clinical isolates from patients
 RT in Rondonia (Brazilian western Amazon region).";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF333312; AAK19583.1; -.
 FT NON_TER 1 1
 FT NON_TER 63 63
 SQ SEQUENCE 63 AA; 7108 MW; ED958F8E1B8B5F7 CRC64;
 Query Match 4.3%; Score 7; DB 5; Length 63;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 QTNNYS 114
 |||||
 DB 13 QTNNYS 19

RESULT 29
 O75702 PRELIMINARY; PRT; 73 AA.
 AC O75702:
 DT 01-NOV-1998 (TREMBlrel. 08; Created)
 DT 01-NOV-1998 (TREMBlrel. 08; Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21; Last annotation update)
 DE Protein-tyrosine-phosphatase, isoform 3 (EC 3.1.3.48).
 GN ACPI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Modesti A., Marzocchi R., Raugel G., Chiti A., Sereni A.,
 RA Madhetti F., Ramponi G.;
 RT "Cloning, expression and characterisation of a new human low Mr
 RT phosphotyrosine protein phosphatase originating by alternative
 RT splicing.";

RL FEBS Lett. 0:0-0(0).
 DR EMBL: Y16846; CAA76416.1; -.
 DR HSSP; P24666; SPNT.
 DR InterPro; IPR000106; Low_mwt_PTPase.
 DR Pfam; PF01451; LMWpc; 1.
 DR SMART; SM00226; LMWpc; 1.
 KW Hydrolase.
 SQ SEQUENCE 73 AA; 7660 MW; D261205427CHEFC3 CRC64;

Query Match 4.3%; Score 7; DB 4; Length 73;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 98 GSLLAVL 104
 |||||
 DB 50 GSLLAVL 56

RESULT 30
 O9ND29 PRELIMINARY; PRT; 74 AA.
 AC O9ND29:
 DT 01-OCT-2000 (TREMBlrel. 15; Created)
 DT 01-OCT-2000 (TREMBlrel. 15; Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19; Last annotation update)
 DE Variant surface protein (Fragment).
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MT08;
 RX MEDLINE=20372615; Pubmed=10910718;
 RA Kirchbatter K., Mosbach R., del Portillo H.A.;
 RT "Plasmodium falciparum: DBL-1 var sequence analysis in field isolates
 RT from central Brazil.";
 RL Exp. Parasitol. 95:154-157(2000).
 DR EMBL; AF172785; AAF89775.1; -.
 FT NON_TER 1 1
 FT NON_TER 74 74
 SQ SEQUENCE 74 AA; 8328 MW; DFA8947BF49A2E61 CRC64;
 Query Match 4.3%; Score 7; DB 5; Length 74;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 QTNNYS 114
 |||||
 DB 16 QTNNYS 22

RESULT 31
 O31846 PRELIMINARY; PRT; 87 AA.
 AC O31846:
 DT 01-JAN-1998 (TREMBlrel. 05; Created)
 DT 01-JAN-1998 (TREMBlrel. 05; Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20; Last annotation update)
 DE YOZN protein.
 GN YOZN.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; Pubmed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borst R., Boutsier L., Brans A., Braum M., Biggell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,

RA Dentzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entlan K.D., Errington J., Fabret C., Ferrari E., Fougier D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizzi A., Galleron N.,
RA Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kiazerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogilwa A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Plescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsstra P., Tognoni A.,
RA Tosato V., Uchlyama S., Vandenbol M., Vannier F., Vassartotl A.,
RA Viari A., Mamult R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zunslein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis.";
RT Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; 299114; CAB13819.1; -.
KW Complete proteome.
SQ SEQUENCE 87 AA; 9671 MW; D36E195DA06C2C24 CRC64;

Query Match 4.3%; Score 7; DB 16; Length 87;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 VNWGONL 9
Db 26 VNWGONL 32

RESULT 32
Q9N3B4 PRELIMINARY; PRT; 107 AA.
AC Q9N3B4;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Hypothetical 11.9 kDa protein.
GN Y54G2A.19.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pseudocercariae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=21016719; PubMed=1130712;
RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressey T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujil C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,

RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC024817; AAF59579.1; -.
KW Hypothetical protein.
SQ SEQUENCE 107 AA; 11887 MW; 5A8456C4FB5B185 CRC64;
Query Match 4.3%; Score 7; DB 5; Length 107;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 VYNSRTD 75
Db 69 VYNSRTD 75

RESULT 33
Q9FE65 PRELIMINARY; PRT; 119 AA.
AC Q9FE65;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Putative 60S ribosomal protein L34 (60S ribosomal protein L34,
DE putative) (T6C23.18) (F24J1.23/F24J1.23).
GN T6C23.18 OR F24J1.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Sakano H., Liu S.X., Pham P.K., Yamada K., Banh J., Etgu P., Lee J.M.,
RA Toriumi M., Yu G., Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene T6C23.18 (GI:665554).";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Bentto M., Creasy T.H., Haas B.J., Wu D.,
RA Maiti R., Romming C.M., Koo H., Fujil C.Y., Uterback T.R.,
RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome 1 BAC F24J1 genomic sequence.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Chao Q., Chen H., Kim C., Ecker J.;
RT "Full length cDNA sequences of Arabidopsis thaliana.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carnilici P., Chen H., Cheuk R., Hayashizaki Y., Ishida Y., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene T6C23.18 (GI:665554).";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=1130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressey T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujil C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,

RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luos J.S., Maili R., Marzalli A.,
 RA Miltcher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant *Arabidopsis*
 RT *thaliana*.";
 RL Nature 408:816-820(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
 RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
 RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Shinozaki K., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA "Arabidopsis cDNA clones.";
 RT Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RA Kim C.J., Chen H., Cheuk R., Meyers M.C., Banh J., Bowser L.,
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.,
 RT "Arabidopsis ORF clones.";
 RT Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AB027531; AAC42912.1; -
 DR EMBL: AC021046; AAC12705.1; -
 DR EMBL: AF024703; AAG40054.1; -
 DR EMBL: AF349526; AAK15573.1; -
 DR EMBL: AC013289; AAG52537.1; -
 DR EMBL: AY052720; AAK96624.1; -
 DR EMBL: AF446885; AAL38618.1; -
 DR InterPro: IPR001284; Ribosomal_L34E.
 DR Pfam: PF01199; Ribosomal_L34E; 1.
 DR PRINTS: PR01250; RIBOSOMAL_L34E.
 DR ProDom: PD005148; Ribosomal_L34E; 1.
 DR PROSITE: PS01145; RIBOSOMAL_L34E; 1.
 KW Ribosomal protein.
 SQ SEQUENCE 119 AA; 13650 MW; 9116C7E50F46E627 CRC64;
 Query Match 4.3%; Score 7; DB 10; Length 119;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 38 AYGCVLS 44
 Db 77 AYGCVLS 83
 RESULT 34
 O9LJM6 PRELIMINARY; PRT; 120 AA.
 AC O9LJM6:
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 60S ribosomal protein L34 (Putative 60S ribosomal protein L34).
 GN AFG68900.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RX SEQUENCE FROM N.A.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.,
 RL Submitted (Jul-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA MEDLINE=20363099; PubMed=10907853;
 RA Nakamura Y.,
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl.
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
 RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
 RA "Arabidopsis Open Reading Frame (ORF) Clones.";
 RT Ecker J.R., Theologis A.,
 RT "Full length cDNA of gene MLD15.7/AF3428900 (GI:9294231).";
 RT Submitted (Jul-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Karlin-Neumann G., Kawai J., Jones T., Kamiya A.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
 RA Davis R.W., Ecker J.R., Theologis A.,
 RT "Arabidopsis Open Reading Frame (ORF) Clones.";
 RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AP000386; BAB02133.1; -
 DR EMBL: AY045818; AAK76492.1; -
 DR EMBL: AY091359; AAM14298.1; -
 DR InterPro: IPR001284; Ribosomal_L34E.
 DR Pfam: PF01199; Ribosomal_L34E; 1.
 DR PRINTS: PR01250; RIBOSOMAL_L34E.
 DR ProDom: PD005148; Ribosomal_L34E; 1.
 DR PROSITE: PS01145; RIBOSOMAL_L34E; 1.
 KW Ribosomal protein.
 SQ SEQUENCE 120 AA; 13650 MW; 2C3D7E7631FDF5A CRC64;
 Query Match 4.3%; Score 7; DB 10; Length 120;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 38 AYGCVLS 44
 Db 77 AYGCVLS 83
 RESULT 35
 Q9RXH5 PRELIMINARY; PRT; 128 AA.
 AC Q9RXH5:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein DR0338.
 GN DR0338.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-R1:
 RA MEDLINE-20036896; PubMed-10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
 RA Makarewicz K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioreistant bacterium Deinococcus
 RL radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001894; AAF09928.1; -.
 DR TIGR: DR0338; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 128 AA; 14083 MW; 63AE3D9F10FEE928 CRC64;

Query Match 4.3%; Score 7; DB 16; Length 128;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 43 LSNFSCT 49
 Db 23 LSNFSCT 29

RESULT 36
 O8U8J1 PRELIMINARY; PRT; 162 AA.
 AC O8U8J1;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein At04100.
 GN At04100 OR AGC_L_1506.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-21608550; PubMed-11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
 RA Chapman P., Clendenning J., Decherage G., Gilliet W., Grant C.,
 RA Kutyavlin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Semplichmachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tilgney S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 RT C58.";
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-21608551; PubMed-11743194;
 RA Goodner B., Hinkle G., Gelling S., Miller N., Blanchard M.,
 RA Querollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmel K., Gordon J., Vaudin M., Iatchouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
 RA Cleto C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58.";
 RL Science 294:2323-2328(2001).
 DR EMBL: AE009341; AAL4901.1; -.
 DR EMBL: AE008274; AAK89331.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 162 AA; 18483 MW; 6362912EABE190B0 CRC64;

Query Match 4.3%; Score 7; DB 16; Length 162;

Best Local Similarity 100.0%; Pred. No. 42;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 4 NVGNLTV 10
 Db 56 NVGNLTV 62

RESULT 37
 O8VV80 PRELIMINARY; PRT; 177 AA.
 AC O8VV80;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE F0F1-ATPase subunit delta.
 GN ATPH.
 OS Colwellia maris (Vibrio sp. (strain ABE-1)).
 OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
 OC Colwellia.
 OX NCBI_TaxID=77524;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Takada Y., Takiya S.;
 RT "Colwellia maris alp operon, complete sequence.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB035129; BAB82481.1; -.
 DR InterPro: IPR000711; ATPsynT_OSCP.
 DR Pfam: PF00213; OSCP.1.
 DR PRINTS: PR00125; ATPASEDELTA.
 DR PROSITE: PS00389; ATPASE_DELTA.
 SQ SEQUENCE 177 AA; 19131 MW; 1CF62614457A0A78 CRC64;

Query Match 4.3%; Score 7; DB 2; Length 177;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 92 GLVIRAG 98
 Db 150 GLVIRAG 156

RESULT 38
 O9U7J9 PRELIMINARY; PRT; 179 AA.
 AC O9U7J9;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Erythrocyte membrane protein 1 SD105E (Fragment).
 GN VAR.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-99404835; PubMed-10477185;
 RA Mard C.P., Clotey G.T., Dorris M., Ji D.D., Arnot D.E.;
 RT "Analysis of Plasmodium falciparum pFEMP-1/var genes suggests that
 RT recombination rearranges constrained sequences.";
 RL Mol. Biochem. Parasitol. 102:167-177(1999).
 DR EMBL: AF127288; AAD52768.1; -.
 FT NON_TER 1 179
 FT NON_TER 1 179
 SQ SEQUENCE 179 AA; 20577 MW; F625DA80FC75086 CRC64;

Query Match 4.3%; Score 7; DB 5; Length 179;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 QTNVNS 114
 Db 8 QTNVNS 14

```

RESULT 39
09P001 PRELIMINARY; PRT; 188 AA.
AC 09P001:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Cytochrome B561.
GN XF1328.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facinchi A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnsels J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasak H.E.,
RA da Silva A.C.F., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.A., Verjovski-Almeida S., Vettore A.L.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL: AE003965; AAF84137.1; -.
KW Complete proteome.
SQ SEQUENCE 188 AA; 21209 MW; 1A22CAFBA0BAEA82 CRC64;

Query Match 4.3%; Score 7; DB 16; Length 188;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 LIAVLIL 106
Db 20 LIAVLIL 26

RESULT 40
09CHD6 PRELIMINARY; PRT; 210 AA.
AC 09CHD6:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein y1ad.
GN Y1AD OR LI0796.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jallion O., Malame K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL: AE006313; AA04894.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 210 AA; 23075 MW; 00025205406FC480 CRC64;

Query Match 4.3%; Score 7; DB 16; Length 210;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 96 KAGSLIA 102
Db 12 KAGSLIA 18

RESULT 41
005437 PRELIMINARY; PRT; 248 AA.
AC 005437:
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein RV3908.
GN RV3908 OR MTCY15F10.03C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Quail M.A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Skelton S., Squares S.,
RA Rutter S., Seeger K., Skelton S., Squares S.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL: Z94121; CAB08093.1; -.
DR TuberculList; RV3908; -.
DR InterPro; IPR000086; NUDIX_hydrolase.
DR Pfam; PF00293; NUDIX; 1.
DR PRINTS; PR00502; NUDIXFAMILY.
DR PROSITE; PS00893; NUDIX; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 248 AA; 27171 MW; 270A5193861EAD03 CRC64;

Query Match 4.3%; Score 7; DB 16; Length 248;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 89 SAGGLVI 95
Db 66 SAGGLVI 72

RESULT 42
08VIR0 PRELIMINARY; PRT; 248 AA.
AC 08VIR0:
DT 01-MAR-2002 (TREMBlrel. 20, Created)

```

DT 01-MAR-2002 (TReMBLrel. 20, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE Muty/mudix family protein.
GN MT4027.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriales; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson M.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,
RA Belcher A., Uitterlind T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AE007193; AAK48392.1; -.
DR TIGR: MT4027; -.
DR InterPro: IPR000086; NUDIX_hydrolase.
DR Pfam: PF00293; NUDIX_1.
DR PRINTS: PR00502; NUDIXFAMILY.
DR PROSITE: PS00893; NUDIX; UNKNOWN_1.
SQ SEQUENCE 248 AA; 27125 MW; 989C08F615D19120 CRC64;

Query Match
Best Local Similarity 4.3%; Score 7; DB 16; Length 248;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 SAGGLVI 95
Db 66 SAGGLVI 72
|||||
PRT; 251 AA.
AC Q50195; PRELIMINARY;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, last annotation update)
DE L222-ORF6 (Hypothetical protein ML2698).
GN ML2698.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriales; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97124199; PubMed=86965312;
RA Fajhl H., De Rosel E., Salazar L., Cantoni R., Labo M., Riccardi G.,
RA Takif H.E., Elgimeier K., Bergh S., Cole S.T.;
RT "Gene arrangement and organization in a approximately 76 kb fragment
RT encompassing the orf6 region of the chromosome of Mycobacterium
RT leprae."
RL Microbiology 142:3147-3161(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Elgimeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,
RA Holtz S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).

DR EMBL: L39923; AAB53126.1; -.
DR EMBL: AL583926; CAC32230.1; -.
DR Lepidoptera; ML2698; -.
DR InterPro: IPR000086; NUDIX_hydrolase.
DR Pfam: PF00293; NUDIX_1.
DR PRINTS: PR00502; NUDIXFAMILY.
DR PROSITE: PS00893; NUDIX; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 251 AA; 28083 MW; 86F4E04D734219F6 CRC64;

Query Match
Best Local Similarity 4.3%; Score 7; DB 16; Length 251;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 SAGGLVI 95
Db 69 SAGGLVI 75
|||||
PRT; 256 AA.
AC Q8R805; PRELIMINARY;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE Dehydrogenases with different specificities (related to short-chain
DE alcohol dehydrogenases).
GN FAB3 OR TFE234.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome."
RL Genome Res. 12:689-700(2002).
DR EMBL: AE013167; AAM25384.1; -.
KW Complete proteome.
SQ SEQUENCE 256 AA; 28153 MW; B08E48F177C61314 CRC64;

Query Match
Best Local Similarity 4.3%; Score 7; DB 16; Length 256;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GONLVVD 12
Db 244 GONLVVD 250
|||||
PRT; 283 AA.
AC Q9ZVF6; PRELIMINARY;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-JUN-2002 (TReMBLrel. 21, last sequence update)
DE Expressed protein (At2g01490/FF219.11).
GN At2G01490.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,

RA Shen M., Romning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Town C.D., Kaul S.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT *Arabidopsis cDNA clones.*
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC005560; AAC67325.2; -;
 DR EMBL: AY057710; AAL15340.1; -;
 SQ SEQUENCE 283 AA; 32028 MW; 1C4520E012ABD1C7 CRC64;

Query Match 4.3%; Score 7; DB 10; Length 283;
 Best local Similarity 100.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 KAGSLIA 102
 DB 225 KAGSLIA 231

Search completed: November 28, 2002, 19:05:33
 Job time : 99 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 19:00:51 ; Search time 19 seconds

(without alignments)
814.613 Million cell updates/sec

Title: us-09-900-575-29_COPY_26_186

Perfect score: 161

Sequence: 1 PNVNCGNLVVDLSTQIFCH.....DVTVLPDYRGSVPIPLRVY 161

Scoring table: OLIGO

Searched: 283224 seqs, 96134422 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database: PIR_73:*

1: plr1:*\n2: plr2:*\n3: plr3:*\n4: plr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	57.1	300	2	S56545
2	90	55.9	300	2	G91288
3	90	55.9	300	2	B86130
4	42	26.1	302	2	A32801
5	5	5.6	304	2	A64904
6	9	5.6	304	2	C90892
7	9	5.6	304	2	F85725
8	5	5.0	860	2	A84833
9	5	5.0	1234	2	T30160
10	8	5.0	1873	2	T30944
11	8	4.3	87	2	B69932
12	7	4.3	119	2	H96717
13	7	4.3	120	2	S48027
14	7	4.3	120	2	S60476
15	7	4.3	120	2	F86395
16	7	4.3	128	2	B75531
17	7	4.3	132	1	MPNR2
18	7	4.3	162	2	AG8226
19	7	4.3	162	2	AG3060
20	7	4.3	188	2	B82696
21	7	4.3	210	2	D86724
22	7	4.3	248	2	E70600
23	7	4.3	251	2	H87246
24	7	4.3	291	2	F69559
25	7	4.3	292	2	S24169
26	7	4.3	298	2	E97096
27	7	4.3	304	2	H75081
28	7	4.3	311	2	T32776
29	7	4.3	311	2	T32776

30	7	4.3	338	2	T38762	hypothetical prote
31	7	4.3	383	2	AD0183	probable exported
32	7	4.3	406	2	AG2704	MFS permease [lipo
33	7	4.3	414	2	G97486	probable mfs trans
34	7	4.3	443	2	B86849	polysaccharide bios
35	7	4.3	498	2	F98149	hypothetical prote
36	7	4.3	533	2	AR3138	hypothetical prote
37	7	4.3	590	2	E75501	sensor histidine k
38	7	4.3	629	2	S29685	retroviral recepto
39	7	4.3	630	2	G84334	threonyl-tRNA synt
40	7	4.3	687	2	S74621	carbon dioxide con
41	7	4.3	741	2	D81798	lactoferrin-bindin
42	7	4.3	761	2	G90728	probable enzyme (l
43	7	4.3	761	2	H85579	probable enzyme yb
44	7	4.3	761	2	C64813	ybhJ protein - Esc
45	7	4.3	790	2	T30081	hypothetical prote
46	7	4.3	839	2	D97013	probably secreted
47	7	4.3	1001	2	C86181	hypothetical prote
48	7	4.3	1004	2	T00795	26S proteasome reg
49	7	4.3	1040	2	B64974	hypothetical prote
50	7	4.3	1040	2	C90889	hypothetical prote
51	7	4.3	1040	2	F85834	hypothetical prote
52	7	4.3	1040	2	AD0771	probable RND-famil
53	7	4.3	1040	2	H83329	probable RND efflu
54	7	4.3	1052	2	AI0346	AcRB/AcrD/AcrE fam
55	7	4.3	1939	2	D97316	probable S-layer P
56	7	3.7	14	2	S57574	T cell receptor V-
57	7	3.7	29	2	S46929	leg169 protein - m
58	6	3.7	43	2	S53436	superoxide dismuta
59	6	3.7	60	2	D82806	hypothetical prote
60	6	3.7	68	1	WMBPBUB	gene c protein - P
61	6	3.7	68	1	JO2005	hypothetical 7.5K
62	6	3.7	73	2	T07913	probable aspartic
63	6	3.7	90	2	T18068	hypothetical prote
64	6	3.7	91	2	F69069	hypothetical prote
65	6	3.7	94	2	F69799	conserved hypotet
66	6	3.7	100	2	F72751	hypothetical prote
67	6	3.7	111	2	S51353	cytochrome-c oxida
68	6	3.7	111	2	E71063	hypothetical prote
69	6	3.7	111	2	E75084	hypothetical prote
70	6	3.7	111	2	A64917	probable membrane
71	6	3.7	111	2	A90918	hypothetical prote
72	6	3.7	111	2	F85766	hypothetical prote
73	6	3.7	111	2	AB0690	probable membrane
74	6	3.7	115	2	B86885	50S ribosomal prot
75	6	3.7	118	2	H81666	conserved hypotet
76	6	3.7	119	2	D97043	hypothetical prote
77	6	3.7	122	2	C37281	replication factor
78	6	3.7	126	2	I60079	opsin, middle-wave
79	6	3.7	128	2	A35690	mucin 3 (clone SIB
80	6	3.7	130	2	G81254	30S ribosomal prot
81	6	3.7	131	2	C97282	ribosomal protein
82	6	3.7	131	2	UC4767	E3 gene encoding h
83	6	3.7	133	2	JC2003	NADH ubiquinone ox
84	6	3.7	134	2	C72341	propionyl-CoA carb
85	6	3.7	136	1	PN0501	phosphoribosyltran
86	6	3.7	137	1	F64961	hypothetical prote
87	6	3.7	137	1	C85815	hypothetical prote
88	6	3.7	137	2	D90967	hypothetical prote
89	6	3.7	139	2	S77060	transposase sl1066
90	6	3.7	144	2	A96580	hypothetical prote
91	6	3.7	145	2	B87560	hypothetical prote
92	6	3.7	147	2	AB1791	hypothetical prote
93	6	3.7	147	2	AC1415	hypothetical prote
94	6	3.7	149	2	S60474	hypothetical prote
95	6	3.7	150	2	JO1183	histone H2A - gard
96	6	3.7	153	2	T05637	hypothetical prote
97	6	3.7	154	2	D97415	hypothetical prote
98	6	3.7	155	2	T22815	hypothetical prote
99	6	3.7	156	2	T23109	hypothetical prote
100	6	3.7	157	2	E83629	hypothetical prote
101	6	3.7	159	2	S42986	probable signal pe
102	6	3.7	160	2	E95882	hypothetical trans

249	6	3.7	330	2	C75192	probable 1-aminoc	322	6	3.7	405	2	B97171	uncharacterized co
250	6	3.7	333	2	AH2038	hypothetical prote	323	6	3.7	406	2	T19887	hypothetical prote
251	6	3.7	335	2	H82046	transcription repr	324	6	3.7	407	2	T18308	60S ribosomal prot
252	6	3.7	336	2	H81091	probable CDP-6-deo	325	6	3.7	407	2	H81160	hypothetical prote
253	6	3.7	336	2	C98230	hypothetical prote	326	6	3.7	408	2	A71657	hypothetical prote
254	6	3.7	338	1	R5H83L	ribosomal protein	327	6	3.7	408	2	B75635	phosphoenolpyruvat
255	6	3.7	338	1	D69405	Na+/H+ antiporter	328	6	3.7	411	2	T34585	probable secreted
256	6	3.7	338	2	T23351	hypothetical prote	329	6	3.7	412	2	D97795	hypothetical prote
257	6	3.7	340	2	S76294	peptide transport	330	6	3.7	412	2	G64059	probable serine tr
258	6	3.7	340	2	A64902	transcription fact	331	6	3.7	412	2	H81415	transcription fact
259	6	3.7	340	2	D38245	transcription fact	332	6	3.7	413	2	S04655	hypothetical prote
260	6	3.7	341	2	A35777	cathepsin D (EC 3.	333	6	3.7	414	2	T16602	hypothetical prote
261	6	3.7	344	1	KHPCD	UDP-N-acetylmuram	334	6	3.7	415	2	G72335	hypothetical prote
262	6	3.7	345	2	B90264	hypothetical prote	335	6	3.7	415	2	T32490	probable C3HC4 zin
263	6	3.7	345	2	AH0457	hypothetical prote	336	6	3.7	416	2	B84710	hypothetical prote
264	6	3.7	345	2	E86202	probable transfe	337	6	3.7	417	2	D72372	quinolone resist
265	6	3.7	346	2	H70722	transcription fact	338	6	3.7	418	2	S74875	probable hydrolase
266	6	3.7	346	2	A38245	transcription fact	339	6	3.7	422	2	A98306	probable Na+ depen
267	6	3.7	347	2	A3306	transcription fact	340	6	3.7	423	2	AB0054	conserved hypot
268	6	3.7	347	2	A39123	probable oligopept	341	6	3.7	425	2	A69770	hypothetical prote
269	6	3.7	349	2	A71212	opsin, 563nm - whi	342	6	3.7	425	2	A90055	probable MoaA prote
270	6	3.7	350	2	D71273	hypothetical prote	343	6	3.7	426	2	E70601	probable MoaA prote
271	6	3.7	350	2	S43497	peptidase, M23/M37	344	6	3.7	426	2	F72402	probable C3HC4 zin
272	6	3.7	350	2	D72259	myb-related protei	345	6	3.7	428	2	S48836	glutathione-disulf
273	6	3.7	351	2	H90178	leucine dehydrogen	346	6	3.7	428	2	T05253	CER2 prote
274	6	3.7	351	2	H87674	conserved hypot	347	6	3.7	428	2	C87286	Xaa-Pro dipeptid
275	6	3.7	352	1	S58293	glutamate 5'-kinase	348	6	3.7	433	2	A71559	dermal glycoprotei
276	6	3.7	353	2	A11859	hypothetical prote	349	6	3.7	434	2	H81709	probable transcrip
277	6	3.7	353	2	D69049	hypothetical prote	350	6	3.7	434	2	T14329	N utilization subs
278	6	3.7	355	2	F70397	hypothetical prote	351	6	3.7	437	2	C35147	glutathione-disulf
279	6	3.7	356	2	A86213	hypothetical prote	352	6	3.7	437	2	JC4988	high-affinity gluc
280	6	3.7	356	2	A98250	hypothetical prote	353	6	3.7	437	2	S18407	integrase homolog
281	6	3.7	356	2	AB0306	opsin, red-sensiti	354	6	3.7	437	2	C87286	hypothetical prote
282	6	3.7	359	2	S40746	opsin, green-sensit	355	6	3.7	437	2	T47831	hypothetical prote
283	6	3.7	364	1	OOHUG	O-sialoglycoprotei	356	6	3.7	438	2	A98161	probable gluconat
284	6	3.7	365	2	AB2902	probable O-sialogl	357	6	3.7	438	2	H86006	hypothetical prote
285	6	3.7	365	2	D97677	O-sialoglycoprotei	358	6	3.7	438	2	AD0481	hypothetical prote
286	6	3.7	367	2	T10207	oleoyl-lacyl-carri	359	6	3.7	438	2	AB0997	hypothetical prote
287	6	3.7	367	2	D72376	pectate lyase - Th	360	6	3.7	438	2	S64418	high-affinity gluc
288	6	3.7	370	2	S69718	hypothetical prote	361	6	3.7	445	2	E85691	hypothetical prote
289	6	3.7	371	2	S12604	collagen alpha 2(V	362	6	3.7	446	2	AF1942	probable tail comp
290	6	3.7	371	2	C90176	thiamin biosynthes	363	6	3.7	446	2	AB2512	hypothetical prote
291	6	3.7	375	2	C71917	probable transamin	364	6	3.7	449	2	C70535	hypothetical prote
292	6	3.7	375	2	H64597	hypothetical prote	365	6	3.7	450	2	AD3117	polygalacturonase
293	6	3.7	375	2	F70576	hypothetical prote	366	6	3.7	452	2	G70905	hypothetical prote
294	6	3.7	375	2	E70576	MID2 protein - yea	367	6	3.7	452	2	F70552	probable glna3 pro
295	6	3.7	376	2	S52137	hypothetical prote	368	6	3.7	452	2	AB2512	hypothetical prote
296	6	3.7	379	2	T04469	conserved hypot	369	6	3.7	455	1	T21089	probable lpgf prot
297	6	3.7	382	2	H90127	casein kinase II (370	6	3.7	455	1	AB2512	replicative DNA he
298	6	3.7	383	1	G69201	casein kinase II (371	6	3.7	455	1	A29939	acid phosphatase (
299	6	3.7	384	2	B30319	GTP-binding protei	372	6	3.7	455	1	T21089	epoxide hydrolase
300	6	3.7	384	2	G86652	RING finger protei	373	6	3.7	455	2	S05573	deoxyribodipyrimid
301	6	3.7	388	2	S49445	uncharacterized co	374	6	3.7	456	2	D70772	hypothetical prote
302	6	3.7	388	2	E97260	casein kinase (EC	375	6	3.7	457	2	F83064	probable MFS trans
303	6	3.7	388	2	E97260	casein kinase II (376	6	3.7	462	1	S36123	phosphotransferase
304	6	3.7	391	2	I49141	casein kinase II (377	6	3.7	462	2	H91104	H+-transporting tw
305	6	3.7	391	2	A30319	casein kinase II (378	6	3.7	462	2	C85950	hypothetical prote
306	6	3.7	391	2	S21335	casein kinase II (379	6	3.7	462	1	PMBSBP	H+-transporting tw
307	6	3.7	391	2	JN0555	casein kinase II (380	6	3.7	467	2	C84420	hypothetical prote
308	6	3.7	392	1	A38611	casein kinase II (381	6	3.7	468	2	T26081	hypothetical prote
309	6	3.7	392	1	B69852	3-ketocacyl-CoA thi	382	6	3.7	469	2	S17726	H+-transporting tw
310	6	3.7	392	2	B69411	hypothetical prote	383	6	3.7	470	2	D84863	conserved hypot
311	6	3.7	394	2	F86190	hypothetical prote	384	6	3.7	473	2	B82765	hypothetical prote
312	6	3.7	394	2	T10200	hypothetical prote	385	6	3.7	476	2	T33273	transcription fact
313	6	3.7	394	2	T20633	hypothetical prote	386	6	3.7	481	2	AH0410	hypothetical prote
314	6	3.7	395	1	RRNZVT	polymerase-associa	387	6	3.7	482	1	EXPR	hypothetical prote
315	6	3.7	395	1	RRNZP2	polymerase-associa	388	6	3.7	482	2	T43996	global stress regu
316	6	3.7	395	2	AB2977	hippurate hydrolas	389	6	3.7	483	2	A55033	coagulation factor
317	6	3.7	395	2	C97244	NADH:flavin oxidor	390	6	3.7	484	2	T44183	keratin 12 - mouse
318	6	3.7	396	2	A13554	xylose transport s	391	6	3.7	484	2	T36538	probable envelope
319	6	3.7	397	1	KITIG	phosphoglycerate k	392	6	3.7	484	2	JC7350	probable transmem
320	6	3.7	397	1	D75387	NADH2 dehydrogenas	393	6	3.7	485	2	D91182	probable outer mem
321	6	3.7	405	2	E81946	hypothetical prote	394	6	3.7	485	2	H86028	probable fibrin

395	3.7	485	2	AE1221	two-component sens	468	6	3.7	610	2	F82192	ABC transporter, A
396	3.7	485	2	AE1574	two-component sens	469	6	3.7	610	2	S05807	SAM1 protein - yea
397	3.7	487	2	T05117	hypothetical prote	470	6	3.7	612	2	T42243	probable polypepti
398	3.7	487	2	T00779	hypothetical prote	471	6	3.7	614	2	T20795	hypothetical prote
399	3.7	491	2	A70681	probable dcta prot	472	6	3.7	615	1	KEF012	coagulation factor
400	3.7	492	2	T38093	probable protein d	473	6	3.7	619	2	164087	translational elonga
401	3.7	493	2	B83071	hypothetical prote	474	6	3.7	623	2	T52656	phosphoglucosylase
402	3.7	495	1	Q1AD52	early E1B 55k prot	475	6	3.7	623	2	A33622	SAC1 protein - yea
403	3.7	495	1	T12401	NADH2 dehydrogenas	476	6	3.7	623	2	T28958	hypothetical prote
404	3.7	496	1	O1AD55	early E1B 55k prot	477	6	3.7	623	2	T23232	hypothetical prote
405	3.7	496	1	H70839	hypothetical glycl	478	6	3.7	636	2	Q0VZ11	early transcriptio
406	3.7	499	2	C71563	probable leucyl am	479	6	3.7	637	1	F72162	P6R protein - vacc
407	3.7	499	2	B75577	beta lactamase-rel	480	6	3.7	637	2	E42515	D6R protein - vacc
408	3.7	500	2	G71255	probable sodium/pr	481	6	3.7	637	2	C36847	early transcriptio
409	3.7	500	2	T26786	hypothetical prote	482	6	3.7	637	2	T28534	hypothetical prote
410	3.7	501	2	B75462	hypothetical prote	483	6	3.7	637	2	T37379	VEP-1, early tran
411	3.7	504	2	T05543	hypothetical prote	484	6	3.7	637	2	JH0611	glutamate-cysteine
412	3.7	508	2	S74848	neopullulanase - S	485	6	3.7	639	2	JU00479	glucan 1,4-alpha-g
413	3.7	509	2	A96749	unknown protein T1	486	6	3.7	639	2	JU00607	glucan 1,4-alpha-g
414	3.7	510	2	H81369	purh bifunctional	487	6	3.7	640	1	ALAS6R	glucan 1,4-alpha-g
415	3.7	512	2	T09330	XIRF1 protein - hu	488	6	3.7	643	2	E82481	glucan 1,4-alpha-g
416	3.7	512	2	AF3402	exopolysphatase	489	6	3.7	648	2	AB3013	metilyl-accepting c
417	3.7	513	2	A87324	hypothetical prote	490	6	3.7	650	2	JC4673	protein kinase (EC
418	3.7	514	2	H70699	probable ppp prote	491	6	3.7	653	2	D69815	conserved hypotet
419	3.7	522	2	T45340	moa protein limpo	492	6	3.7	656	2	AE1479	probable cell surf
420	3.7	523	2	C70717	probable purh prot	493	6	3.7	662	2	E98271	metalloproteinase
421	3.7	524	2	T07988	ATP-dependent clp	494	6	3.7	664	1	TNBE70	5k alpha trans-
422	3.7	525	2	G84406	halolysin limporte	495	6	3.7	667	2	S64915	EMF70 protein prec
423	3.7	525	2	E83913	hypothetical prote	496	6	3.7	670	2	T02019	callus-associated
424	3.7	527	2	T45439	probable phosphori	497	6	3.7	671	2	S51599	Om(2D) protein - f
425	3.7	530	2	T16972	probable high affi	498	6	3.7	672	2	I40333	tracheal colonizat
426	3.7	531	2	T20763	hypothetical prote	499	6	3.7	674	2	D81220	NADH2 dehydrogenas
427	3.7	533	2	H86749	hypothetical prote	500	6	3.7	674	2	P81990	hypothetical prote
428	3.7	539	2	S67049	probable membrane	501	6	3.7	681	2	T39814	conserved hypotet
429	3.7	541	1	W2EPPY	pectate lyase (EC	502	6	3.7	681	2	T39814	erythrocyte membra
430	3.7	545	1	D90159	hypothetical prote	503	6	3.7	685	2	B69401	75.8k alpha trans-
431	3.7	548	2	G82286	phosphate ABC tran	504	6	3.7	691			

541	6	3.7	820	2	H86246	614	6	3.7	1363	2	C84346	hypothetical prote
542	6	3.7	825	2	T23612	615	6	3.7	1379	2	JC5778	apoptosis signal-r
543	6	3.7	834	2	SS4563	616	6	3.7	1387	2	A96771	hypothetical prote
544	6	3.7	835	2	F90260	617	6	3.7	1390	2	T31353	hypothetical prote
545	6	3.7	839	2	A56337	618	6	3.7	1391	2	S73652	RNA polymerase bet
546	6	3.7	840	2	T02164	619	6	3.7	1398	2	C87448	hypothetical prote
547	6	3.7	847	2	C87678	620	6	3.7	1400	1	T81818	protein-tyrosine k
548	6	3.7	849	2	C90834	621	6	3.7	1405	2	H87230	probable integral
549	6	3.7	850	2	T23306	622	6	3.7	1407	1	T00558	P-glycoprotein-11k
550	6	3.7	853	1	SS4553	623	6	3.7	1417	2	T47671	hypothetical prote
551	6	3.7	859	2	TRBPHL	624	6	3.7	1440	2	T44872	probable integral
552	6	3.7	869	2	SS3098	625	6	3.7	1451	2	S41035	hypothetical prote
553	6	3.7	878	2	A83748	626	6	3.7	1458	2	S36014	dyslin heavy chain
554	6	3.7	879	2	T02728	627	6	3.7	1473	2	T13855	suppressor of smbl
555	6	3.7	905	2	F82734	628	6	3.7	1492	2	T18560	DNA-directed DNA p
556	6	3.7	910	2	H82826	629	6	3.7	1494	2	T13798	hypothetical prote
557	6	3.7	917	2	S09646	630	6	3.7	1500	2	AF2027	hypothetical prote
558	6	3.7	920	2	I40614	631	6	3.7	1508	2	T27828	probable DNA-direc
559	6	3.7	921	2	A48184	632	6	3.7	1513	2	T28158	hypothetical prote
560	6	3.7	922	2	A45183	633	6	3.7	1519	2	T27829	proliferation pote
561	6	3.7	925	2	T06128	634	6	3.7	1550	2	T42727	hypothetical prote
562	6	3.7	930	2	D86546	635	6	3.7	1560	2	T42727	hypothetical prote
563	6	3.7	930	2	D72078	636	6	3.7	1563	2	T01879	hypothetical prote
564	6	3.7	931	2	A81591	637	6	3.7	1633	2	B91052	hypothetical prote
565	6	3.7	934	2	T43725	638	6	3.7	1653	2	F85896	hypothetical prote
566	6	3.7	936	2	S75633	639	6	3.7	1653	2	G65028	genome polypeptid
567	6	3.7	936	2	E71405	640	6	3.7	1691	1	A44212	ataxia telangiecta
568	6	3.7	942	2	D96814	641	6	3.7	1708	1	A43100	collagen type IV a
569	6	3.7	948	2	A57640	642	6	3.7	1761	2	T13990	hypothetical prote
570	6	3.7	951	2	T00017	643	6	3.7	1771	2	T13990	collagen alpha 1(X
571	6	3.7	951	2	F97442	644	6	3.7	1774	2	B56101	hypothetical prote
572	6	3.7	967	2	AH2660	645	6	3.7	1857	2	TDHUK	leukoocyte antigen
573	6	3.7	969	2	S17886	646	6	3.7	1897	1	TDHUK	hypothetical prote
574	6	3.7	978	2	T16948	647	6	3.7	2136	2	A05037	scavenger receptor
575	6	3.7	989	2	A82140	648	6	3.7	2136	2	A05037	hemagglutinin/hemo
576	6	3.7	997	2	S33754	649	6	3.7	2153	2	T09083	cell surface anti
577	6	3.7	1003	2	H82883	650	6	3.7	2153	2	T09083	nonstructural poly
578	6	3.7	1010	2	T09499	651	6	3.7	2240	2	B71704	inositol 1,4,5-tri
579	6	3.7	1014	2	S75724	652	6	3.7	2240	2	B71704	cyclic beta 1-2 gl
580	6	3.7	1017	2	T30195	653	6	3.7	2514	1	MMWV2	probable membrane
581	6	3.7	1017	2	P82500	654	6	3.7	2783	2	T31431	hypothetical prote
582	6	3.7	1018	1	CGH02A	655	6	3.7	2870	2	H96974	BLF1 protein - hu
583	6	3.7	1026	1	T18220	656	6	3.7	3013	2	AB0480	hypothetical prote
584	6	3.7	1029	1	S21369	657	6	3.7	3029	2	S76109	hypothetical prote
585	6	3.7	1034	2	S63536	658	6	3.7	3149	1	Q08B8	hypothetical prote
586	6	3.7	1035	2	T07836	659	6	3.7	3449	2	D86161	hypothetical prote
587	6	3.7	1037	2	A36096	660	6	3.7	3600	2	S28600	MEGF1 protein - ra
588	6	3.7	1053	2	S46199	661	6	3.7	4085	2	T00252	dyslin heavy chain
589	6	3.7	1057	2	T04874	662	6	3.7	4351	2	A38905	hypothetical prote
590	6	3.7	1057	2	A42109	663	6	3.7	4644	1	A44357	hypothetical prote
591	6	3.7	1060	2	S33641	664	6	3.7	4725	1	A44357	hypothetical prote
592	6	3.7	1081	2	S15040	665	6	3.7	4936	2	AH2515	hypothetical prote
593	6	3.7	1086	2	AH2136	666	6	3.7	5005	2	F82884	hypothetical prote
594	6	3.7	1097	1	S47220	667	6	3.7	5232	2	A45086	HC-toxin synthetas
595	6	3.7	1102	2	S35617	668	6	3.1	14	2	PT0223	pax-QNR, long form
596	6	3.7	1121	2	C82120	669	6	3.1	14	2	PT0223	Ig heavy chain CDR
597	6	3.7	1124	2	T30340	670	6	3.1	15	2	P00778	NADH2 dehydrogenas
598	6	3.7	1127	2	F84570	671	6	3.1	18	2	A28027	protein p2 - curie
599	6	3.7	1131	2	T04587	672	6	3.1	24	2	A41037	antimicrobial pept
600	6	3.7	1134	2	T04587	673	6	3.1	25	2	A41037	neurotrophin u - com
601	6	3.7	1139	2	A40932	674	6	3.1	25	2	C69350	hypothetical prote
602	6	3.7	1166	2	T28680	675	6	3.1	29	2	A49410	hypothetical prote
603	6	3.7	1169	2	T28680	676	6	3.1	29	2	B85840	hypothetical prote
604	6	3.7	1174	2	A40853	677	6	3.1	29	2	B85840	hypothetical prote
605	6	3.7	1185	2	T19212	678	6	3.1	29	2	A18780	hypothetical prote
606	6	3.7	1203	2	T04294	679	6	3.1	32	2	B85607	hypothetical prote
607	6	3.7	1203	2	T21275	680	6	3.1	32	2	B85607	hypothetical prote
608	6	3.7	1203	2	T21275	681	6	3.1	32	2	B85607	hypothetical prote
609	6	3.7	1240	2	T03097	682	6	3.1	32	2	A64848	hypothetical prote
610	6	3.7	1251	2	T21389	683	6	3.1	35	2	E38601	Ig kappa chain V r
611	6	3.7	1270	2	T51227	684	6	3.1	37	2	A81552	hypothetical prote
612	6	3.7	1290	2	A56493	685	6	3.1	38	2	S78357	photosystem II pro
613	6	3.7	1315	2	A56101	686	6	3.1	43	2	C72270	gene MLV protein -

Tue Dec 3 12:14:29 2002

us-09-900-575-29_copy_26_186.oligo.rpr

687	5	3.1	44	2	B89984	hypothetical prote	760	77	2	H69328	nifu protein (nifu
688	5	3.1	45	2	I39208	H+-exporting ATPase	761	77	2	C69902	permease homolog y
689	5	3.1	46	2	S20018	hypothetical prote	762	77	2	A37317	probable membrane
690	5	3.1	47	2	G87679	hypothetical prote	763	78	2	A27683	heat shock 90K pro
691	5	3.1	48	2	S13571	hypothetical prote	764	79	1	T1EPVK	K+ channel blocker
692	5	3.1	49	2	S29173	D-galactose-bindin	765	80	2	AB2803	conserved hypothet
693	5	3.1	50	2	F82335	hypothetical prote	766	80	2	AE3392	phosphoribosylform
694	5	3.1	51	2	AC2221	hypothetical prote	767	80	2	EB4630	probable small nuc
695	5	3.1	51	2	I40723	lysc protein - Cor	768	80	2	S19984	hypothetical prote
696	5	3.1	51	2	E82981	50S ribosomal prot	769	80	2	S19987	hypothetical prote
697	5	3.1	51	2	E48652	Integral membrane	770	80	2	F98090	hypothetical prote
698	5	3.1	51	2	T36214	spdc protein - Str	771	81	2	S41580	lysosome (RC 3.2.1
699	5	3.1	52	2	A10472	hypothetical prote	772	81	2	S22826	small nuclear ribo
700	5	3.1	52	2	A11345	hypothetical prote	773	81	2	T35731	transport protein
701	5	3.1	53	2	E95311	hypothetical prote	774	81	2	T47865	mitochondrial prot
702	5	3.1	53	2	R5EC33	hypothetical prote	775	81	2	A24522	hypothetical prote
703	5	3.1	55	1	E82347	ribosomal protein	776	81	2	AB3462	hypothetical prote
704	5	3.1	55	2	AD0007	50S ribosomal prot	777	81	2	AF2553	probable transport
705	5	3.1	55	2	AF0971	50S ribosomal chai	778	81	2	AF2553	hypothetical prote
706	5	3.1	55	2	G19192	50S ribosomal subu	779	82	2	C95126	hypothetical prote
707	5	3.1	55	2	H86039	interferon alpha r	780	83	2	S70360	hypothetical prote
708	5	3.1	56	2	S41602	hypothetical prote	781	83	2	C69440	hypothetical prote
709	5	3.1	57	2	AD0065	ferredoxin [import	782	83	2	H85523	hypothetical prote
710	5	3.1	58	2	C84363	protein F0766.5 [i	783	83	2	AD0081	hypothetical prote
711	5	3.1	60	2	E89468	translation elonga	784	83	2	AB2036	hypothetical prote
712	5	3.1	61	2	B61510	carbon storage reg	785	84	2	E86209	hypothetical prote
713	5	3.1	61	2	D83531	hypothetical prote	786	84	2	T17676	hypothetical prote
714	5	3.1	61	2	C81531	conserved hypothet	787	84	2	C39247	hypothetical prote
715	5	3.1	63	2	G96531	hypothetical prote	788	85	2	E39247	hypothetical prote
716	5	3.1	63	2	E84370	hypothetical prote	789	85	2	F30536	hypothetical prote
717	5	3.1	63	2	T37351	RNA polymerase sub	790	85	2	F30536	hypothetical prote
718	5	3.1	63	2	T28506	hypothetical prote	791	85	2	I68800	hypothetical prote
719	5	3.1	63	2	B72159	I6R protein - Vari	792	85	2	D39247	hypothetical prote
720	5	3.1	64	2	C87511	hypothetical prote	793	85	2	A39247	hypothetical prote
721	5	3.1	64	2	B72740	probable ribonucle	794	85	2	I68795	hypothetical prote
722	5	3.1	64	2	AF3568	hypothetical prote	795	85	2	I68795	hypothetical prote
723	5	3.1	65	2	E84494	hypothetical prote	796	85	2	I68795	hypothetical prote
724	5	3.1	65	2	E97099	hypothetical prote	797	85	2	I68796	hypothetical prote
725	5	3.1	66	2	F90146	DNA-directed RNA p	798	85	2	I68802	hypothetical prote
726	5	3.1	66	2	G98059	hypothetical prote	800	85	2	I68799	hypothetical prote
727	5	3.1	67	1	O7BPE7	gene 17.5 protein	801	85	2	T03332	hypothetical prote
728	5	3.1	67	1	O7BPE3	lysine protein - ph	802	86	2	D33989	hypothetical prote
729	5	3.1	67	1	G62568	cold shock protein	803	86	2	F34964	hypothetical prote
730	5	3.1	67	2	T42055	hypothetical prote	804	86	2	D64371	hypothetical prote
731	5	3.1	67	2	T07561	hypothetical prote	805	86	2	G30536	hypothetical prote
732	5	3.1	67	2	A71289	probable DNA-direc	806	87	2	I59648	hypothetical prote
733	5	3.1	68	2	AH3199	hypothetical prote	807	87	2	I59648	hypothetical prote
734	5	3.1	68	2	T44551	hypothetical prote	808	87	2	S15699	hypothetical prote
735	5	3.1	68	2	AD2807	hypothetical prote	809	87	2	T43570	hypothetical prote
736	5	3.1	69	2	B83524	hypothetical prote	810	87	2	A71426	hypothetical prote
737	5	3.1	69	2	B83374	hypothetical prote	811	87	2	F40361	hypothetical prote
738	5	3.1	69	2	B69335	hypothetical prote	812	87	2	C97909	hypothetical prote
739	5	3.1	69	2	S41000	hypothetical prote	813	87	2	W4WL18	hypothetical prote
740	5	3.1	69	2	H82753	hypothetical prote	814	87	2	D90957	hypothetical prote
741	5	3.1	70	2	S35030	photosystem I chai	815	88	1	H66752	pathogen-induced p
742	5	3.1	70	2	I39058	hypothetical prote	816	88	2	T06988	hypothetical prote
743	5	3.1	70	2	C97412	hypothetical prote	817	88	2	G83514	hypothetical prote
744	5	3.1	71	2	I51747	MHC class II alpha	818	88	2	S29977	hypothetical prote
745	5	3.1	71	2	S74348	hypothetical prote	819	88	2	AB7563	hypothetical prote
746	5	3.1	71	2	AF0683	hypothetical prote	820	89	2	F34768	hypothetical prote
747	5	3.1	71	2	AI2240	hypothetical prote	821	90	2	S45092	hypothetical prote
748	5	3.1	72	2	D87340	hypothetical prote	822	90	2	T15357	hypothetical prote
749	5	3.1	72	2	G83106	hypothetical prote	823	90	2	T09295	hypothetical prote
750	5	3.1	72	2	E83733	hypothetical prote	824	90	2	F70630	hypothetical prote
751	5	3.1	73	2	D86616	hypothetical prote	825	90	2	C90667	hypothetical prote
752	5	3.1	73	2	T44614	hypothetical prote	826	91	2		
753	5	3.1	73	2	C72008	hypothetical prote	827	91	2		
754	5	3.1	74	2	AD3290	protein translocat	828	91	2		
755	5	3.1	76	2	C84094	hypothetical prote	829	91	2		
756	5	3.1	76	2	T17960	metallothionein -	830	91	2		
757	5	3.1	76	2	S52636	auxin-induced prot	831	92	2		
758	5	3.1	77	2	T05712		832	92	2		

833	3.1	92	2	A90764	hypothetical prote	906	5	3.1	105	2	D85677	unknown protein en
834	3.1	92	2	A90819	hypothetical prote	907	5	3.1	105	2	H80817	hypothetical prote
835	3.1	92	2	S22829	copr protein - Str	908	5	3.1	105	2	D97507	hypothetical prote
836	3.1	92	2	G85517	hypothetical prote	909	5	3.1	105	2	A13472	hypothetical prote
837	3.1	93	1	WMH89	3c protein - avian	910	5	3.1	105	2	H71635	hypothetical prote
838	3.1	93	2	T10379	hypothetical prote	911	5	3.1	105	2	C97846	hypothetical prote
839	3.1	93	2	AE1247	hypothetical prote	912	5	3.1	106	1	WMH23	3c protein - avian
840	3.1	93	2	A11609	hypothetical prote	913	5	3.1	106	2	S28688	hypothetical prote
841	3.1	94	2	C96576	hypothetical prote	914	5	3.1	106	2	B11659	hypothetical prote
842	3.1	94	2	JH0207	hypothetical 10.8K	915	5	3.1	106	2	T24774	hypothetical prote
843	3.1	94	2	T50619	hypothetical prote	916	5	3.1	106	2	C70794	hypothetical prote
844	3.1	94	2	AG1365	hypothetical prote	917	5	3.1	106	2	AB0256	hypothetical prote
845	3.1	95	2	S30068	polysialacturonase	918	5	3.1	106	2	G69477	probable piage pro
846	3.1	95	2	S30062	polysialacturonase	919	5	3.1	107	1	WMH26	NADH2 dehydrogenas
847	3.1	95	2	G37262	Ig heavy chain V r	920	5	3.1	107	2	S77775	3c protein - avian
848	3.1	95	2	A41872	heat shock protein	921	5	3.1	107	2	168730	probable CPD1acyl
849	3.1	95	2	G97232	co-chaperonin GroE	922	5	3.1	107	2	168726	IgE chain C3 regio
850	3.1	95	2	B4216	hypothetical prote	923	5	3.1	107	2	AG3490	hypothetical trans
851	3.1	95	2	T45374	hypothetical prote	924	5	3.1	107	2	E71634	glutaredoxin-like
852	3.1	96	2	D47624	Ig heavy chain V-I	925	5	3.1	107	2	G83788	hypothetical prote
853	3.1	96	2	E86773	citrate (pro-3S)-1	926	5	3.1	107	2	B71069	hypothetical prote
854	3.1	96	2	G86801	prophage p13 prote	927	5	3.1	108	1	S09879	hypothetical prote
855	3.1	96	2	F81890	hypothetical prote	928	5	3.1	108	1	H64913	hypothetical prote
856	3.1	96	2	S72376	pds6 protein - Ent	929	5	3.1	108	2	C87494	probable membrane
857	3.1	96	2	E69063	hypothetical prote	930	5	3.1	108	2	H83992	hypothetical prote
858	3.1	97	2	S75342	hypothetical prote	931	5	3.1	108	2	H83992	hypothetical prote
859	3.1	97	2	T45331	hypothetical prote	932	5	3.1	108	2	E85763	conserved hypotet
860	3.1	97	2	T81655	hypothetical prote	933	5	3.1	108	2	F90033	probable membrane
861	3.1	97	2	C97072	uncharacterized pr	934	5	3.1	108	2	AH0679	F911.29 protein -
862	3.1	98	2	AB1133	hypothetical prote	935	5	3.1	108	2	H86287	hypothetical prote
863	3.1	98	2	UC5822	NADH2 dehydrogenas	936	5	3.1	108	2	A12333	hypothetical prote
864	3.1	99	2	S25971	hypothetical prote	937	5	3.1	108	2	F64545	hypothetical prote
865	3.1	99	2	T01678	hypothetical prote	938	5	3.1	108	2	D72470	hypothetical prote
866	3.1	100	1	BVMY7B	chaperonin groES -	939	5	3.1	109	1	OOIH11	3c protein - avian
867	3.1	100	1	BMVY7B	chaperonin groES -	940	5	3.1	109	1	B69806	conserved hypotet
868	3.1	100	2	A64076	urease (EC 3.5.1.5	941	5	3.1	109	2	AD2091	anti-sigma factor
869	3.1	100	2	S25180	heat shock protein	942	5	3.1	110	1	R5BY22	acidic ribosom
870	3.1	100	2	S62755	cerato-ulinm precu	943	5	3.1	110	1	K65Y22	hypothetical prote
871	3.1	100	2	A64016	hypothetical prote	944	5	3.1	110	1	D64341	hypothetical prote
872	3.1	100	2	C72375	hypothetical prote	945	5	3.1	110	2	F71095	hypothetical prote
873	3.1	101	1	PFH04	platelet factor 4	946	5	3.1	110	2	A72661	hypothetical prote
874	3.1	101	1	T38824	thioltransferase -	947	5	3.1	110	2	H64512	hypothetical prote
875	3.1	101	2	D75013	hypothetical prote	948	5	3.1	110	2	G70021	hypothetical prote
876	3.1	101	2	H95941	hypothetical prote	949	5	3.1	110	2	T50423	hypothetical prote
877	3.1	101	2	D96576	hypothetical prote	950	5	3.1	110	2	A38529	niKa protein - Bsc
878	3.1	101	2	S30975	stage V sporulatio	951	5	3.1	110	2	S75767	hypothetical prote
879	3.1	101	2	JQ1300	calgizazarin - Rabb	952	5	3.1	110	2	B69403	hypothetical prote
880	3.1	102	2	S37565	chaperonin groES -	953	5	3.1	110	2	D82628	hypothetical prote
881	3.1	102	2	A41325	heat shock protein	954	5	3.1	110	2	UJ0120	hypothetical prote
882	3.1	102	2	C87282	conserved hypotet	955	5	3.1	111	1	I48640	hypothetical prote
883	3.1	102	2	F69475	conserved hypotet	956	5	3.1	111	2	B45403	ribonuclease-relat
884	3.1	102	2	H82648	conserved hypotet	957	5	3.1	111	2	G11839	insulin-like growt
885	3.1	102	2	K4RB	Ig kappa-Bk chain	958	5	3.1	111	2	G71288	ribosomal protein
886	3.1	103	1	G64093	hemolysin A - Prot	959	5	3.1	111	2	D97843	conserved hypotet
887	3.1	103	1	S04101	conserved hypotet	960	5	3.1	111	2	G34768	glutaredoxin-like
888	3.1	103	2	B81160	phosphoribosyl-ATP	961	5	3.1	111	2	S64364	hypothetical prote
889	3.1	103	2	S54839	probable membrane	962	5	3.1	111	2	B81106	probable membrane
890	3.1	103	2	S64330	B26r protein - vac	963	5	3.1	111	2	S33180	hypothetical prote
891	3.1	103	2	G42528	hypothetical prote	964	5	3.1	112	2	A86604	nitrogen regulator
892	3.1	103	2	B70600	platelet factor 4	965	5	3.1	112	2	B72019	periplasmic divale
893	3.1	103	2	A71817	probable thiodox	966	5	3.1	112	2	F70954	probable 18r2 prot
894	3.1	104	2	B64702	thiodioxin - Heli	967	5	3.1	112	2	B43601	LSR2 T-cell antige
895	3.1	104	2	F53275	Ig kappa-1 chain C	968	5	3.1	112	2	S66512	pediocin productio
896	3.1	104	2	B69058	calgizazarin - huma	969	5	3.1	113	2	S73951	ribosomal protein
897	3.1	105	1	I37080	ribosomal protein	970	5	3.1	113	2	F72687	hypothetical prote
898	3.1	105	1	C71635	50S ribosomal prot	971	5	3.1	113	2	AF3643	hypothetical prote
899	3.1	105	2	B97845	platelet factor 4	972	5	3.1	114	2	P00745	phycochrome - Meso
900	3.1	105	2	A26774	chortion protein -	973	5	3.1	114	2	T30866	hypothetical prote
901	3.1	105	2	S23061	hypothetical prote	974	5	3.1	114	2	AF2315	hypothetical prote
902	3.1	105	2	F70730		975	5	3.1	114	2	S75565	hypothetical prote
903	3.1	105	2			976	5	3.1	114	2	H96541	hypothetical prote
904	3.1	105	2			977	5	3.1	114	2		
905	3.1	105	2			978	5	3.1	114	2		

```

979      5      3.1      115      2      B32071      T-cell receptor de
980      5      3.1      115      2      S16534      metallochionein-11
981      5      3.1      115      2      F84469      Probable glycine-r
982      5      3.1      115      2      T20559      hypothetical prote
983      5      3.1      115      2      E83184      hypothetical prote
984      5      3.1      116      1      S34611      nitrogenase (Ec 1.
985      5      3.1      116      2      S35468      NMDH2 dehydrogenas
986      5      3.1      116      2      S39434      spermadhesin AON-3
987      5      3.1      116      2      S17567      hypothetical prote
988      5      3.1      116      2      F64510      hypothetical prote
989      5      3.1      116      2      T15004      conserved hypocher
990      5      3.1      116      2      F82805      hypothetical prote
991      5      3.1      116      2      D75355      hypothetical prote
992      5      3.1      116      2      A98319      hypothetical prote
993      5      3.1      116      2      A02964      coat protein - rab
994      5      3.1      117      1      VCWWRH      quinone oxidoreduc
995      5      3.1      117      2      T41247      probable lipid tra
996      5      3.1      117      2      T02872      developmental regu
997      5      3.1      117      2      A45387      ribosomal protein
998      5      3.1      117      2      C64116      ribosomal protein
999      5      3.1      117      2      B82476      coat protein - rab
1000     5      3.1      117      2      S22135

```

ALIGNMENTS

```

RESULT 1
S56545
fimbrial protein flmH precursor, type 1 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 01-Mar-2002
C:Accession: S56545; B65246; S09563; A36967
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
A:Reference number: S56314; MUID:9334362; PMID:7610040
A:Accession: S56545
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-300 <BLAT>
A:Cross-references: EMBL:U14003; NID:91263172; PIDD:AAA97216.1; PID:9537161
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
-A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:9742617; PMID:9278503
A:Accession: B65246
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-300 <BLAT>
A:Cross-references: GB:AE000502; GB:U00096; NID:92367374; PIDD:AACT7276.1; PID:91790775;
R:Klemm, P.; Christiansen, G.
Mol. Gen. Genet. 208, 439-445, 1987
A:Title: Three flm genes required for the regulation of length and mediation of adhesion
A:Reference number: S07321; MUID:86038337; PMID:2890081
A:Accession: S09563
A:Molecule type: DNA
A:Residues: 1-196, 'R', 198-221, 'H', 223-300 <KLE>
A:Cross-references: EMBL:X05672; NID:941463; PIDD:CAA29156.1; PID:941466
R:Sokurenko, E.V.; Courtney, H.S.; Ohman, D.E.; Klemm, P.; Hasty, D.L.
J. Bacteriol. 176, 748-755, 1994
A:Title: FlmH family of type I fimbrial adhesins: functional heterogeneity due to minor
A:Reference number: A36967; MUID:94131954; PMID:7905476
A:Accession: A36967
A:Status: not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-78, 'R', 80-221, 'H', 223-300 <SOK>
A:Note: sequence extracted from NCBI backbone (NCBIP:143314)
C:Genetics:
A:Gene: flmH

```

```

C:Function:
A:Description: involved in longitudinal regulation and mannose-specific adhesion
A:Note: not necessary for the production of fimbriae
A:Note: controls length and number of fimbriae
C:Superfamily: fimbrial protein flmH
C:Keywords: fimbria
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-300/Product: fimbrial protein flmH, type 1 #status predicted <MAT>

```

```

Query Match      57.1%; Score 92; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 5.6e-87;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1  PYNVGNLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLSNFGTVKSGSSYPPT 60
      |||
Db      47  PYNVGNLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLSNFGTVKSGSSYPPT 106
      |||
QY      61  TTSETPRVYNSRDKPVPALYLPVSSAG 92
      |||
Db      107 TTSETPRVYNSRDKPVPALYLPVSSAG 138
      |||

```

RESULT 2

```

G91288
hypothetical protein ECs5279 [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: G91288
R:Hayashi, T.; Makino, K.; Ohishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhrara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G91288
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-300 <HAY>
A:Cross-references: GB:BA000007; PIDD:BA838702.1; PID:91364757; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs5279
C:Superfamily: fimbrial protein flmH
Query Match      55.9%; Score 90; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 6.5e-85;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      3  YNVGNLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLSNFGTVKSGSSYPPT 62
      |||
Db      49  YNVGNLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLSNFGTVKSGSSYPPT 108
      |||
QY      63  SETPRVYNSRDKPVPALYLPVSSAG 92
      |||
Db      109 SETPRVYNSRDKPVPALYLPVSSAG 138
      |||

```

RESULT 3

```

B86130
hypothetical protein flmH [imported] - Escherichia coli (strain O157:H7, substrain ED
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B86130
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
ller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074955; PMID:11206551
A:Accession: B86130
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-300 <SNO>
A:Cross-references: GB:AE005174; NID:912519327; PIDD:AA659502.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933

```

C:Genetics:
A:Gene: flmH
C:Superfamily: fimbrial protein flmH

Query Match
Best Local Similarity 55.9%; Score 90; DB 2; Length 300;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VNNGNLVVDLSTQIFCHNDYPERITDYVTLQSGAVGLSMFGSTVXSGSSYFPPTT 62
Db 49 VNNGNLVVDLSTQIFCHNDYPERITDYVTLQSGAVGLSMFGSTVXSGSSYFPPTT 108
Y 63 SEPRVYVNSRTDKPMPVALYLRPVSSAG 92
Db 109 SEPRVYVNSRTDKPMPVALYLRPVSSAG 138

RESULT 4

A32801
fimbrial adhesin precursor, type 1 - Klebsiella pneumoniae
C:Species: Klebsiella pneumoniae
C:Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 18-Sep-1998
C:Accession: A32801
R:Gerlach, G.F.; Clegg, S.; Allen, B.L.
J. Bacteriol. 171, 1262-1270, 1989
A:Title: Identification and characterization of the genes encoding the type 3 and type 1
A:Reference number: A32801; PMID:8915420; PMID:2563996
A:Accession: A32801
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-302 <GER>
A:Cross-references: GB:M24564
C:Superfamily: fimbrial protein flmH

Query Match
Best Local Similarity 26.1%; Score 42; DB 2; Length 302;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VNNGNLVVDLSTQIFCHNDYPERITDYVTLQSGAVGLSMFGSTVXSGSSYFPPTT 44
Db 51 VNNGNLVVDLSTQIFCHNDYPERITDYVTLQSGAVGLSMFGSTVXSGSSYFPPTT 92

RESULT 5

A64904
probable fimbrial protein b1502 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: A64904
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co-
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64904; PMID:9278503
A:Accession: A64904
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-304 <BLAT>
A:Cross-references: GB:AE000247; GB:U00096; NID:G1787773; PIDN:AAC74575.1; PID:G1787779;
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: fimbrial protein flmH
C:Keywords: fimbria

Query Match
Best Local Similarity 5.6%; Score 9; DB 2; Length 304;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GONLVVDLS 14
Db 53 GONLVVDLS 61

RESULT 6

C90892
probable adhesin [imported] - Escherichia coli (strain O157:H7, substrain RMD 050995)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: C90892
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gasawara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A96029; PMID:21156231; PMID:11258796
A:Accession: C90892

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-304 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA035530.1; PID:G13361573; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs2107
C:Superfamily: fimbrial protein flmH

Query Match
Best Local Similarity 5.6%; Score 9; DB 2; Length 304;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GONLVVDLS 14
Db 53 GONLVVDLS 61

RESULT 7

F85725
probable adhesin, flm type protein Z2206 [imported] - Escherichia coli (strain O157:
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F85725
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May-
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda-
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; PMID:21074935; PMID:11206551
A:Accession: F85725
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-304 <STO>
A:Cross-references: GB:AE005174; NID:G12515169; PIDN:AAG56266.1; GSPDB:GN00145; UNGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2206
C:Superfamily: fimbrial protein flmH

Query Match
Best Local Similarity 5.6%; Score 9; DB 2; Length 304;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GONLVVDLS 14
Db 53 GONLVVDLS 61

RESULT 8

F75497
probable arginine utilization protein RocoB - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: F75497
R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; PMID:20036896; PMID:10567266
A:Accession: F75497
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-539 <WHI>
A:Cross-references: GB:AE001919; GB:AE000513; NID:96458307; PIDN:AAF10189.1; PID:9645830
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0612
C:Superfamily: 1
C:Superfamily: Deinococcus radiodurans probable arginine utilization protein R06B

Query Match
Best Local Similarity 5.0%; Score 8; DB 2; Length 539;
100.0%; Pred. No. 5.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 DV5ARDVT 144
|||||||
Db 4 DV5ARDVT 11

RESULT 9
A84833
hypothetical protein At2g40720 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84833
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84833
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-860 <STO>
A:Cross-references: GB:AE002093; NID:94895221; PIDN:AAD32807.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g40720
A:Map position: 2

Query Match
Best Local Similarity 5.0%; Score 8; DB 2; Length 860;
100.0%; Pred. No. 8.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 VSARDVT 145
|||||||
Db 128 VSARDVT 135

RESULT 10
T30160
hypothetical protein C37A2.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T30160
R:Le, T.T.; Kemp, K.; Scheet, P.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid C37A2.
A:Reference number: Z20746
A:Accession: T30160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1234 <LEF>
A:Cross-references: EMBL:U97194; PIDN:AA852447.1; GSPDB:GN00019; CESP:C37A2.2
A:Experimental source: strain Bristol N2; clone C37A2
C:Genetics:
A:Gene: CESP:C37A2.2
A:Map position: 1
A:Introns: 46/1; 124/1; 159/1; 231/1; 508/3; 740/3; 891/1; 931/1; 975/1; 1014/3; 1083/3

Query Match
Best Local Similarity 5.0%; Score 8; DB 2; Length 1234;
100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 RGSVP1PL 158
|||||||
Db 960 RGSVP1PL 967

RESULT 11
T30944
surface protein precursor - Enterococcus faecalis
C:Species: Enterococcus faecalis
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30944
R:Shankar, V.; Baghdadyan, A.S.; Huycke, M.M.; Lindahl, G.; Gilmore, M.S.
Infect. Immun. 67, 193-200, 1999
A:Title: Infection derived Enterococcus faecalis strains are enriched in esp, a gene
A:Reference number: Z20943; MUID:99081742; PMID:9864215
A:Accession: T30944
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1873 <SHA>
A:Cross-references: EMBL:AF034779; NID:93873186; PID:93873187; PIDN:AAD09858.1

Query Match
Best Local Similarity 5.0%; Score 8; DB 2; Length 1873;
100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 DVTYTLPD 149
|||||||
Db 1754 DVTYTLPD 1761

RESULT 12
B69932
hypothetical protein yozN - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: B69932
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
leeh, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koehler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Takakoshi, A.; Tanaka, T.; Terstira, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbato, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: B69932
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-87 <KUN>
A:Cross-references: GB:Z99114; GB:AL009126; NID:92634230; PIDN:CAB13819.1; PID:el1853
A:Experimental source: strain 168
C:Genetics:
A:Gene: yozN

Query Match
Best Local Similarity 4.3%; Score 7; DB 2; Length 87;
100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VNVGQNL 9
|||||||
Db 26 VNVGQNL 32

RESULT 13
H96717
probable 60S ribosomal protein L34 T6C23.18 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C:Accession: H96717
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzella, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: H96717
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-119 <STO>
 A:Cross-references: GB:AE005173; NID:9665554; PIDN:AAF22923.1; GSPDB:GN00141
 A:Gene: T6C23.18
 A:Map position: 1
 C:Superfamily: rat ribosomal protein L34

Query Match 4.3%; Score 7; DB 2; Length 119;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 38 AYGCVLS 44
 Db 77 AYGCVLS 83

RESULT 14
 S48027
 ribosomal protein L34, cytosolic - common tobacco
 C:Species: Nicotiana tabacum (common tobacco)
 C:Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 13-Aug-1999
 C:Accession: S48027; S48028
 R:Gao, J.; Kim, S.R.; Chung, Y.Y.; Lee, J.M.; An, G.
 A:Title: Developmental and environmental regulation of two ribosomal protein genes in tobacco.
 A:Reference number: S48026; MUID:94355650; PMID:8075394
 A:Accession: S48027
 A:Molecule type: mRNA
 A:Residues: 1-120 <GAO>
 A:Cross-references: EMBL:L27089; NID:9436029; PIDN:AA57158.1; PID:9436030
 A:Experimental source: TSC40-3
 A:Accession: S48028
 A:Molecule type: mRNA
 A:Residues: 1-120 <GA2>
 A:Cross-references: EMBL:L27107; NID:9436031; PIDN:AA57159.1; PID:9436032
 A:Experimental source: TSC40-4
 C:Superfamily: rat ribosomal protein L34
 C:Keywords: protein biosynthesis; ribosome

Query Match 4.3%; Score 7; DB 2; Length 120;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 38 AYGCVLS 44
 Db 77 AYGCVLS 83

RESULT 15
 S60476
 ribosomal protein L34, cytosolic - garden pea
 C:Species: Pisum sativum (garden pea)
 C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 13-Aug-1999
 C:Accession: S60476
 R:Devitt, M.L.; Statistom, J.P.
 A:Title: Sequence and analysis of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: H96717
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-119 <STO>
 A:Cross-references: GB:AE005173; NID:9665554; PIDN:AAF22923.1; GSPDB:GN00141
 A:Gene: T6C23.18
 A:Map position: 1
 C:Superfamily: rat ribosomal protein L34

A:Title: Cell cycle regulation during growth-dormancy cycles in pea axillary buds.
 A:Reference number: S60474; MUID:96046745; PMID:7379177
 A:Accession: S60476
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-120 <DEV>
 A:Cross-references: EMBL:U10047; NID:9498907; PIDN:AA86953.1; PID:9498908
 A:Gene: rpl34
 C:Superfamily: rat ribosomal protein L34
 C:Keywords: protein biosynthesis; ribosome

Query Match 4.3%; Score 7; DB 2; Length 120;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 38 AYGCVLS 44
 Db 77 AYGCVLS 83

RESULT 16
 F86395
 60s ribosomal protein L34 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: F86395
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzella, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: F86395
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-120 <STO>
 A:Cross-references: GB:AE005172; NID:94262177; PIDN:AA014494.1; GSPDB:GN00141
 A:Gene: T6C23.18
 A:Map position: 1
 C:Superfamily: rat ribosomal protein L34

Query Match 4.3%; Score 7; DB 2; Length 120;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 38 AYGCVLS 44
 Db 77 AYGCVLS 83

RESULT 17
 B75531
 hypothetical protein - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
 C:Accession: B75531
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: B75531
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-128 <WHI>
 A:Cross-references: GB:AE001894; GB:AE000513; NID:96458011; PIDN:AAF09238.1; PID:9645

A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0338
 A:Map position: 1
 C:Superfamily: Deinococcus radiodurans hypothetical protein DR0338

Query Match 4.3%; Score 7; DB 2; Length 128;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 43 LSNFSGT 49
 |||||||
 Db 23 LSNFSGT 29

RESULT 18

myelin P2 protein [validated] - human
 N:Alternate names: peripheral myelin protein 2
 C:Species: Homo sapiens (man)
 C>Date: 05-Apr-1983 #sequence_revision 27-Jan-1995 #text_change 08-Dec-2000
 C:Accession: J09777; A03143; S24224
 R:Hayasaka, K.; Nanao, K.; Tahara, M.; Sato, W.; Takada, G.; Miura, M.; Uyemura, K.
 Biochem. Biophys. Res. Commun. 181, 204-207, 1991
 A:Title: Isolation and sequence determination of cDNA encoding P2 protein of human periph
 A:Reference number: J09777; M01D:92068191; PMID:1720307
 A:Accession: J09777
 A:Molecule type: mRNA
 A:Residues: 1-132 <NAY>
 A:Cross-references: EMBL:X62167; NID:935185; PIDN:CAAA4096.1; PID:935186
 A:Experimental source: fetal spinal cord peripheral myelin
 R>Note: authors did not translate the codon for residue 1
 R:Stuuk, M.; Kitamura, K.; Sakamoto, Y.; Uyemura, K.
 J. Neurochem. 39, 1759-1762, 1982
 A:Title: The complete amino acid sequence of human P2 protein.
 A:Reference number: A03143; M01D:83058785; PMID:6183401
 A:Accession: A03143
 A:Molecule type: Protein
 A:Residues: 2-98, 'N', 100-110, 'D', 112-132 <SUZ>
 C:Comment: P2 protein and myelin basic protein together constitute a major fraction of F

C:Genetics:
 A:Gene: GDB:PMP2
 A:Cross-references: GDB:129030; OMIM:170715
 A:Map position: 8q21.3-8q22.1
 C:Superfamily: myelin P2 protein
 C:Keywords: acetylated amino end; myelin; phosphoprotein
 F:2-132/product: myelin P2 protein #status experimental <MAT>
 F:2/Modified site: acetylated amino end (Ser) (in mature form)
 F:20/Binding site: phosphate (Tyr) (covalent) #status predicted
 F:118-125/Disulfide bonds: #status experimental

Query Match 4.3%; Score 7; DB 1; Length 132;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 VTLQGS 37
 |||||||
 Db 85 VTLQGS 91

RESULT 19

A98226
 hypothetical protein AGR_L1506 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C:Species: Agrobacterium tumefaciens
 C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
 C:Accession: A98226
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollan, C.; Allinger, M.; Doughly, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tun
 A:Reference number: A97359; PMID:11743194
 A:Accession: A98226
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-162 <KUR>
 A:Cross-references: GB:AE007870; PIDN:AAK89331.1; PID:915159173; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGR_L1506
 A:Map position: linear chromosome

Query Match 4.3%; Score 7; DB 2; Length 162;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 NVGONLV 10
 |||||||
 Db 56 NVGONLV 62

RESULT 20

AG3060
 conserved hypothetical protein Atu4100 [imported] - Agrobacterium tumefaciens (strain
 C:Species: Agrobacterium tumefaciens
 C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 C:Accession: AG3060
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kulyavin, T.; Levy, R.; Li, M.; McCl
 Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AG3060
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-162 <KUR>
 A:Cross-references: GB:AE008689; PIDN:AAL44901.1; PID:917742552; GSPDB:GN00187
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu4100
 A:Map position: linear chromosome

Query Match 4.3%; Score 7; DB 2; Length 162;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 NVGONLV 10
 |||||||
 Db 56 NVGONLV 62

RESULT 21

B82696
 cytochrome B561 XF1328 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: B82696
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: AB2515; M01D:20365717; PMID:10910347
 A>Note: for a complete list of authors see reference number A59328 below
 A:Accession: B82696
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-188 <STM>
 A:Cross-references: GB:AE003965; GB:AE003849; NID:99106313; PIDN:AAE84137.1; GSPDB:GN
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
 as-Neto, E.; Docena, C.; El-Dorfy, H.; Facinanci, A.P.; Ferreira, A.J.S.
 submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
 J.D.; Jungueira, M.L.; Kemper, E.L.; Kiteajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
Rodrigues, V.; Rosa, A.J. de M.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silva
M.; Tsubako, M.H.; Valada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
C:Contents: annotation
C:Genetics:
A:Gene: XFI328
C:Superfamily: cytochrome b561

Query Match 4.3%; Score 7; DB 2; Length 188;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 LIAVLIL 106
|||||||
DB 20 LIAVLIL 26

RESULT 22

hypothetical protein yIAD [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: D86724

R:Colin, A.; Winkler, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss

A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: D86724

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-210 <STO>

A:Cross-references: GB:AEO05176; PID:912723717; PIDN:AKK04894.1; GSPDB:GN00146

A:Experimental source: strain IL1403

A:Genetics:

Query Match

Best Local Similarity 4.3%; Score 7; DB 2; Length 210;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 KAGSLIA 102
|||||||
DB 12 KAGSLIA 18

RESULT 23

hypothetical protein RV3908 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 26-May-2000
C:Accession: E70600

R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
Rajandream, M.A.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitham, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:9825987; PMID:9634230

A:Accession: E70600

A:Status: preliminary

A:Molecule type: nucleic acid sequence not shown; translation not shown

A:Residues: 1-248 <COL>

A:Cross-references: GB:294121; GB:AL123456; NID:93261736; PIDN:CAB08093.1; PID:e312269;

A:Experimental source: strain H37RV

A:Genetics:

A:Gene: RV3908

C:Superfamily: unassigned mult domain proteins; mult domain homology

F:98-115/Domain: mult domain homology <MULT>

Query Match

4.3%; Score 7; DB 2; Length 248;

Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 89 SAGGLVI 95
|||||||
DB 66 SAGGLVI 72

RESULT 24

H87246

conserved hypothetical protein ML2698 [imported] - Mycobacterium lepre

C:Species: Mycobacterium lepre

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 24-Aug-2001
C:Accession: H87246

R:Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;
eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.

A:Title: Massive gene decay in the leprosy bacillus

A:Reference number: A86909; MUID:21128732; PMID:11234002

A:Accession: H87246

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-251 <STO>

A:Cross-references: GB:AL450380; NID:913093888; PIDN:CAC32230.1; GSPDB:GN00147

A:Genetics:

A:Gene: ML2698

C:Superfamily: unassigned mult domain proteins; mult domain homology

Query Match

Best Local Similarity 4.3%; Score 7; DB 2; Length 251;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 SAGGLVI 95
|||||||
DB 69 SAGGLVI 75

RESULT 25

hypothetical protein ygho - Bacillus subtilis

C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: F69959

R:Kunst, F.; Ogatawa, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabel, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Gallazzi, A.; Gal
Koch, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Huilo, M

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y. M.; Ogawa, K.; Ogatawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

A:Authors: Schleich, S.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scari
A:Authors: Schleich, S.; Schreiber, R.; Scoffone, F.; Seiglitz, J.; Sekowska, A.; Se

A:Authors: Winkler, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: F69959

A:Status: preliminary

A:Molecule type: nucleic acid sequence not shown; translation not shown

A:Residues: 1-291 <RUN>

A:Cross-references: GB:299116; GB:AL009126; NID:92634723; PIDN:CAB14382.1; PID:e11857

A:Experimental source: strain 168

A:Genetics:

A:Gene: ygho

C:Superfamily: unassigned mult domain proteins; mult domain homology

Query Match

Best Local Similarity 4.3%; Score 7; DB 2; Length 291;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 40 GCVLSNF 46
 |||||
 DB 184 GCVLSNF 190

RESULT 26

S24169

mucin - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 10-Nov-1995

C:Accession: S24169

R:huan, L.J.; Xu, G.; Forstner, G.; Forstner, J.

Biochim. Biophys. Acta 1132, 79-82, 1992

A:Title: A serine, threonine and proline-rich region near the carboxyl-terminus of a rat

A:Reference number: S24169

A:Accession: S24169

A:Molecule type: mRNA

A:Residues: 1-292 <HUA>

Query Match 4.3%; Score 7; DB 2; Length 292;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 PTTSETP 66
 |||||
 DB 7 PTTSETP 13

RESULT 27

E97096

zn-binding lipoprotein related (surface adhesin A), ADHS [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 01-Mar-2002

C:Accession: E97096

R:Noelling, J.; Brelton, G.; Omeletchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium CLO

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: E97096

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-298 <KOR>

A:Cross-references: GB:AE001437; PIDN:AAK79560.1; PID:q15024548; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC1593

C:Superfamily: adhesin B

Query Match 4.3%; Score 7; DB 2; Length 298;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 99 SLIAVLI 105
 |||||
 DB 7 SLIAVLI 13

RESULT 28

H75081

phosphate abc transporter, permease protein (pstC) PAB0698 - Pyrococcus abyssi (strain C

C:Species: Pyrococcus abyssi

C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: H75081

R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A:Reference number: A75001

A:Accession: H75081

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-304 <KAW>

A:Cross-references: GB:A1248286; GB:AL096836; NID:q5458366; PIDN:CA849957.1; PID:q545
 A:Experimental source: strain Orsay
 C:Genetics:
 A:Gene: pstC; PAB0698
 C:Superfamily: phoW protein

Query Match 4.3%; Score 7; DB 2; Length 304;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 99 SLIAVLI 105
 |||||
 DB 72 SLIAVLI 78

RESULT 29

T32776

hypothetical protein D1069.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000

C:Accession: T32776

R:Murray, J.; Langston, Y.; Clarke, K.; Morris, M.

submitted to the EMBL Data Library, December 1997

A:Description: The sequence of C. elegans cosmid D1069.

A:Reference number: Z21223

A:Accession: T32776

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-311 <MUR>

A:Cross-references: EMBL:AF040641; PIDN:AA894946.1; GSPDB:GN00020; CESP:D1069.1

A:Experimental source: strain Bristol N2; clone D1069

C:Genetics:

A:Gene: CESP:D1069.1

A:Map position: 2

A:Insertions: 264/3

C:Superfamily: Caenorhabditis elegans hypothetical protein D1069.1

Query Match 4.3%; Score 7; DB 2; Length 311;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 VVDLSTQ 16
 |||||
 DB 211 VVDLSTQ 217

RESULT 30

T38762

hypothetical protein SPAC3H8.04 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000

C:Accession: T38762

R:Gentles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, January 1996

A:Reference number: Z21810

A:Accession: T38762

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-338 <GEN>

A:Cross-references: EMBL:Z69086; PIDN:CAA93161.1; GSPDB:GN00066; SPDB:SPAC3H8.04

A:Experimental source: strain 972h-; cosmid G3H8

C:Genetics:

A:Gene: SPDB:SPAC3H8.04

A:Map position: 1

C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPAC3H8.04

Query Match 4.3%; Score 7; DB 2; Length 338;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 104 LILROTN 110
 |||||
 DB 237 LILROTN 243

RESULT 31

AD0183
 probable exported protein YPO1504 [Imported] - Yersinia pestis (strain C092)
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
 C:Accession: AD0183
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titchall, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Terraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell, Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MWID:21470413; PMID:11586360
 A:Accession: AD0183
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-383 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC90327.1; PID:g15979546; GSPDB:GN00175
 C:Genetics:
 A:Gene: YPO1504

Query Match
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 90 AGGLVIR 96
 |||||
 Db 366 AGGLVIR 372

RESULT 32

AG2704
 MCS permease [Imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 02-Aug-2002
 C:Accession: AG2704
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; Mclell Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AG2704
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-406 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AL42053.1; PID:g17739431; GSPDB:GN00186
 C:Genetics:
 A:Experimental source: strain C58 (Dupont)
 A:Gene: Atu1040
 A:Map position: circular chromosome
 C:Superfamily: hypothetical protein c0103

Query Match
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 97 AGSLIIV 103
 |||||
 Db 18 AGSLIIV 24

RESULT 33

G97486
 Probable mfs transporter (PA1286) [Imported] - Agrobacterium tumefaciens (strain C58, Ce
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 02-Aug-2002
 C:Accession: G97486
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
 A:Reference number: A97359; PMID:11743194
 A:Accession: G97486
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-414 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK86848.1; PID:g15156062; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C_1918
 A:Map position: circular chromosome
 C:Superfamily: hypothetical protein c0103

Query Match
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 97 AGSLIIV 103
 |||||
 Db 26 AGSLIIV 32

RESULT 34

B86849
 Polysaccharide biosynthesis protein [Imported] - Lactococcus lactis subsp. lactis (str
 C:Species: Lactococcus lactis subsp. lactis
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 C:Accession: B86849
 R:Boletín, A.; Wincker, P.; Manger, S.; Jallón, O.; Malarme, K.; Weissbach, J.; Eh Genome Res. 11, 731-753, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
 A:Reference number: AB6625; MWID:21235186; PMID:11337471
 A:Accession: B86849
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-443 <STO>
 A:Cross-references: GB:AE005176; PID:g12724819; PIDN:AAK05892.1; GSPDB:GN00146
 C:Genetics:
 A:Experimental source: strain IL1403
 A:Gene: ysfC

Query Match
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 89 SAGGLVI 95
 |||||
 Db 57 SAGGLVI 63

RESULT 35

P98149
 Hypothetical protein AGR_L_305 [Imported] - Agrobacterium tumefaciens (strain C58, Ce
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
 C:Accession: P98149
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Gold Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
 A:Reference number: A97359; PMID:11743194
 A:Accession: P98149
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-498 <KUR>
 A:Cross-references: GB:AE007870; PIDN:AAK88720.1; PID:g15158457; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGR_L_305
 A:Map position: linear chromosome
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

Query Match
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 DVTVTLTP 148
|||||||

Db 285 DVTVTLTP 291

RESULT 36

AE3138
hypothetical protein Atu4729 [Imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AE3138
R:Moore, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCell
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AE3138
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-533 <KUR>
A:Cross-references: GB:AE008689; PIDN:AA145523.1; PID:g17743234; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
A:Genetics:
A:Gene: Atu4729
A:Map position: linear chromosome
C:Superfamily: unassigned ATP-binding cassette proteins: ATP-binding cassette homology

Query Match 4.3%; Score 7; DB 2; Length 533;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 DVTVTLTP 148
|||||||

Db 320 DVTVTLTP 326

RESULT 37
F75501
sensor histidine kinase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75501
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M. Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: F75501
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-590 <WHI>
A:Cross-references: GB:AE001917; GB:AE000513; NID:g6458281; PIDN:AAF10157.1; PID:g645827
A:Experimental source: strain R1
A:Genetics:
A:Gene: DR0577
A:Map position: 1

Query Match 4.3%; Score 7; DB 2; Length 590;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 VSSAGGL 93
|||||||

Db 549 VSSAGGL 555

RESULT 38

S29685
retroviral receptor - human

C:Species: Homo sapiens (man)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Nov-2000
C:Accession: S29685; A40775
R:Abritton, M.; Tseng, L.; Cunningham, J.M.
submitted to the EMBL Data Library, January 1991
A:Description: Nucleotide sequence of the human gene similar to the murine ecotropic
A:Reference number: S29685
A:Accession: S29685
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-629 <ABR>
A:Cross-references: EMBL:X57303; NID:g35919; PIDN:CAA40560.1; PID:g35920
R:Yoshimoto, T.; Yoshimoto, E.; Meruelo, D.
Virology 185, 10-17, 1991
A>Title: Molecular cloning and characterization of a novel human gene homologous to t
A:Reference number: A40775; MUID:92024065; PMID:1718082
A:Accession: A40775
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-22, 'R', 24-629 <YOS>
A:Cross-references: GB:X59155; NID:g36160; PIDN:CAA41869.1; PID:g36161
A:Genetics:
A:Gene: REC1L
C:Superfamily: ecotropic retrovirus receptor protein

Query Match 4.3%; Score 7; DB 2; Length 629;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 SLIAYLI 105
|||||||

Db 501 SLIAYLI 507

RESULT 39
G84334
threonyl-tRNA synthetase [Imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: G84334
R:Ng, W.V.; Kennedy, S.P.; Mahairs, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky
R.; Leitner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
A>Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: G84334
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-650 <STO>
A:Cross-references: GB:AE004437; NID:g10561281; PIDN:AAG20043.1; GSPDB:GN00138
A:Genetics:
A:Gene: thrs
C:Superfamily: threonine-tRNA ligase

Query Match 4.3%; Score 7; DB 2; Length 650;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 VTVTLTP 149
|||||||

Db 14 VTVTLTP 20

RESULT 40
S74621
carbon dioxide concentrating mechanism protein comm - Synechocystis sp. (strain PCC 6
N:Alternate names: hypothetical protein sl11031
C:Species: Synechocystis sp.
A:Variety: PCC 6803

C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S74621
R: Kaneo, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M. Res. 8, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74621
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-687 <KAN>
A:Cross-references: EMBL:D90900; GB:AB001339; NID:g1651768; PIDN:BA16773.1; PID:d101750
C:Genetics:
A:Gene: ccm
A:Start codon: GTG

Query Match
Best Local Similarity 4.3%; Score 7; DB 2; Length 687;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 86 PVSSAGG 92
Db 466 PVSSAGG 472

RESULT 41
D81798
Lactoferrin-binding protein NMA1740 [Imported] - *Neisseria meningitidis* (strain Z2491 se
C:Species: *Neisseria meningitidis*
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: D81798
R: Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: D81798
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-741 <PAR>
A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB84968.1; PID:g738038
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: lbpB; NMA1740

Query Match
Best Local Similarity 4.3%; Score 7; DB 2; Length 741;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 48 GTVYSG 54
Db 194 GTVYSG 200

RESULT 42
G90728
Probable enzyme [Imported] - *Escherichia coli* (strain O157:H7, substrain RMD 0509952)
C:Species: *Escherichia coli*
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 24-Aug-2001
C:Accession: G90728
R: Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90728
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-761 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA834222.1; PID:g13360258; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952

C:Genetics:
A:Gene: ECs0799
C:Superfamily: Iron-responsive element-binding protein

Query Match
Best Local Similarity 4.3%; Score 7; DB 2; Length 761;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 95 IKAGSLI 101
Db 747 IKAGSLI 753

RESULT 43
H85579
Probable enzyme ybhJ [Imported] - *Escherichia coli* (strain O157:H7, substrain EDJ933)
C:Species: *Escherichia coli*
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85579
R: Ferna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lm, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85579
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-761 <STO>
A:Cross-references: GB:AE005174; NID:g12513706; PIDN:AGC55100.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDJ933
C:Genetics:
A:Gene: ybhJ
C:Superfamily: Iron-responsive element-binding protein

Query Match
Best Local Similarity 4.3%; Score 7; DB 2; Length 761;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 95 IKAGSLI 101
Db 747 IKAGSLI 753

RESULT 44
C64813
YbhJ protein - *Escherichia coli* (strain K-12)
C:Species: *Escherichia coli*
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: C64813
R: Blatter, F.R.; Plunkett III, G.; Bloch, C.A.; Ferna, N.T.; Burland, V.; Riley, M.;
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: C64813
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-761 <BLAT>
A:Cross-references: GB:AE000179; GB:U00096; NID:g1786978; PIDN:AC73858.1; PID:g17869
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ybhJ
C:Superfamily: Iron-responsive element-binding protein
C:Keywords: transmembrane protein
F:413-429/Domain: transmembrane #status predicted <TM>

Query Match
Best Local Similarity 4.3%; Score 7; DB 2; Length 761;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 95 IKAGSLI 101
Db 747 IKAGSLI 753

RESULT 45

T30081

hypothetical protein C31H2.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000

C:Accession: T30081

R:Geisler, C.; Gatlung, S.

submitted to the EMBL Data Library, November 1995

A:Description: The sequence of C. elegans cosmid C31H2.

A:Reference number: 220732

A:Accession: T30081

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-790 <GET>

A:Cross-references: EMBL:U41748; PIDN:AAA83335.1; CESP:C31H2.1

C:Genetics:

A:Gene: CESP:C31H2.1

A:Introns: 45/3; 78/2; 110/3; 148/3; 198/1; 244/3; 274/3; 318/3; 366/3; 396/3; 427/3; 45

C:Superfamily: Caenorhabditis elegans hypothetical protein C31H2.1

Query Match

4.3%; Score 7; DB 2; Length 790;

Best Local Similarity 100.0%; Pred. No. 83;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 GGYLSNF 46

|||||

DB 210 GGYLSNF 216

Search completed: November 28, 2002, 19:06:08
Job time : 32 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 19:01:31 ; Search time 17 Seconds

(without alignments)
278.652 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186

Perfect score: 161

Sequence: 1 PVTWNGQNLVLDLSTQIFCH.....DVTWTLPRGSRVPLTVY 161

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 segs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	4.3	132	1	US-08-409-731A-11
2	7	4.3	132	2	US-08-470-298B-11
3	7	4.3	132	2	US-09-023-073A-11
4	7	4.3	132	2	US-09-361-737-11
5	7	4.3	629	2	US-08-132-990A-8
6	7	4.3	629	5	PCT-US92-09382-8
7	6	3.7	92	4	US-09-134-001C-4014
8	6	3.7	124	1	US-08-326-117B-13
9	6	3.7	124	3	US-08-982-129-13
10	6	3.7	124	4	US-08-982-129-13
11	6	3.7	145	4	US-09-134-001C-4330
12	6	3.7	185	3	US-08-985-526-36
13	6	3.7	191	4	US-09-561-500-13
14	6	3.7	191	4	US-09-561-500-13
15	6	3.7	191	4	US-09-561-500-13
16	6	3.7	191	4	US-09-561-500-13
17	6	3.7	213	1	US-08-159-784-2
18	6	3.7	213	1	US-08-248-839C-94
19	6	3.7	243	2	US-07-885-089B-4
20	6	3.7	243	2	US-07-885-089B-4
21	6	3.7	243	2	US-07-885-089B-4
22	6	3.7	243	2	US-08-944-483-70
23	6	3.7	243	6	5202428-2
24	6	3.7	249	2	US-08-956-267A-2
25	6	3.7	252	4	US-08-989-386-5
26	6	3.7	252	4	US-09-238-303-10
27	6	3.7	281	4	US-08-481-944-75
28	6	3.7	293	4	US-09-025-769B-178
29	6	3.7	293	4	US-09-512-342-14
30	6	3.7	297	1	US-08-118-270-57
31	6	3.7	297	1	US-08-118-270-58
32	6	3.7	297	5	PCT-US93-08528-57
33	6	3.7	297	5	PCT-US93-08528-58
34	6	3.7	329	4	US-09-134-001C-3238
35	6	3.7	347	2	US-08-424-349A-8
36	6	3.7	347	3	US-08-164-292B-13
37	6	3.7	347	3	US-08-845-623-13
38	6	3.7	347	3	US-08-815-927-13
39	6	3.7	347	4	US-09-103-330-13
40	6	3.7	358	4	US-09-435-242-13
41	6	3.7	366	4	US-09-615-192A-38
42	6	3.7	376	4	US-09-134-001C-4365
43	6	3.7	387	3	US-08-689-421-23
44	6	3.7	387	4	US-09-389-528-23
45	6	3.7	387	4	US-09-181-827A-23
46	6	3.7	391	4	US-09-131-028A-4
47	6	3.7	391	4	US-09-131-028A-4
48	6	3.7	417	5	US-08-867-611-20
49	6	3.7	417	5	PCT-US92-0655A-25
50	6	3.7	445	2	US-08-630-118A-2
51	6	3.7	445	2	US-08-630-118A-2
52	6	3.7	445	2	US-08-838-399-2
53	6	3.7	445	2	US-08-838-399-2
54	6	3.7	445	2	US-09-003-199-2
55	6	3.7	445	2	US-09-003-199-2
56	6	3.7	445	2	US-09-235-839-2
57	6	3.7	445	4	US-09-235-839-2
58	6	3.7	445	4	US-09-327-035-2
59	6	3.7	445	4	US-09-327-035-2
60	6	3.7	447	4	US-09-370-253-6
61	6	3.7	455	3	US-08-993-359-12
62	6	3.7	455	5	PCT-US95-08565-14
63	6	3.7	455	5	US-08-349-025-2
64	6	3.7	456	1	US-08-356-096A-2
65	6	3.7	456	2	US-08-668-650B-2
66	6	3.7	456	4	US-09-200-673-2
67	6	3.7	456	5	PCT-US95-15646-2
68	6	3.7	466	4	US-09-134-001C-3355
69	6	3.7	487	4	US-09-525-046-2
70	6	3.7	516	3	US-08-689-421-29
71	6	3.7	516	4	US-09-389-528-29
72	6	3.7	516	4	US-09-181-827A-29
73	6	3.7	521	3	US-08-996-338-20
74	6	3.7	530	3	US-09-222-817-12
75	6	3.7	530	3	US-09-222-817-12
76	6	3.7	530	3	US-09-222-817-12
77	6	3.7	530	4	US-09-222-786-14
78	6	3.7	531	3	US-08-688-988-39
79	6	3.7	541	1	US-08-604-333-2
80	6	3.7	541	3	US-09-110-618-2
81	6	3.7	541	4	US-09-173-151A-28
82	6	3.7	541	4	US-09-578-178-2
83	6	3.7	551	4	US-09-130-491-16
84	6	3.7	551	4	US-08-221-891-13
85	6	3.7	576	1	US-08-454-439-13
86	6	3.7	576	1	US-08-454-439-13
87	6	3.7	576	4	US-08-464-954A-6
88	6	3.7	579	5	PCT-US94-10487-13
89	6	3.7	579	1	US-08-126-564A-31
90	6	3.7	596	4	PCT-US94-09143-31
91	6	3.7	632	1	US-09-171-337A-6
92	6	3.7	632	1	US-08-221-817-11
93	6	3.7	632	1	US-08-454-439-11
94	6	3.7	632	5	PCT-US94-10487-11
95	6	3.7	637	4	US-09-221-294-4
96	6	3.7	640	4	US-09-351-814-13
97	6	3.7	667	4	US-08-959-004-11
98	6	3.7	675	4	US-08-973-273-5
99	6	3.7	726	4	US-09-413-814-10
100	6	3.7	808	2	US-08-658-335B-33

977 5 3.1 340 4 US-08-975-762-54 Sequence 54, Appl
978 5 3.1 340 4 US-09-295-028-54 Sequence 54, Appl
979 5 3.1 340 4 US-09-106-582-54 Sequence 54, Appl
980 5 3.1 341 1 US-07-748-783-4 Sequence 4, Appl
981 5 3.1 341 1 US-08-166-818-4 Sequence 4, Appl
982 5 3.1 341 4 US-09-134-001C-3783 Sequence 3783, Ap
983 5 3.1 342 2 US-08-689-011-2 Sequence 2, Appl
984 5 3.1 342 2 US-09-373-958-2 Sequence 2, Appl
985 5 3.1 342 4 US-09-381-810A-1 Sequence 2, Appl
986 5 3.1 342 4 US-09-129-033-2 Sequence 2, Appl
987 5 3.1 342 4 US-09-134-001C-5198 Sequence 5198, Ap
988 5 3.1 343 4 US-08-853-948B-3 Sequence 3, Appl
989 5 3.1 344 2 US-07-857-224B-92 Sequence 92, Appl
990 5 3.1 344 2 US-07-857-224B-93 Sequence 93, Appl
991 5 3.1 344 4 US-08-818-112-69 Sequence 70, Appl
992 5 3.1 344 4 US-08-818-111-70 Sequence 70, Appl
993 5 3.1 344 4 US-09-056-556-69 Sequence 69, Appl
994 5 3.1 344 4 US-09-134-001C-4304 Sequence 4304, Ap
995 5 3.1 344 4 US-09-072-596-70 Sequence 70, Appl
996 5 3.1 345 3 US-08-858-003-34 Sequence 34, Appl
997 5 3.1 345 3 US-09-078-166-34 Sequence 34, Appl
998 5 3.1 346 4 US-08-997-467-34 Sequence 34, Appl
999 5 3.1 346 4 US-09-199-657A-313 Sequence 313, App
1000 5 3.1 348 3 US-08-875-540-13 Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-08-409-731A-11
; Sequence 11, Application US/08409731A

Patent No. 5658758
GENERAL INFORMATION:
APPLICANT: NI, Jian
APPLICANT: Yu, Guo-Liang
APPLICANT: Gentz, Reiner
APPLICANT: Rosen, Craig A.
TITLE OF INVENTION: CYTOSTATIN I
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUMAN GENOME SCIENCES, INC.
STREET: 9410 KEY WEST AVENUE
CITY: ROCKVILLE
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,731A
FILING DATE: 24-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Robert H.
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PFI175
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8512
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-409-731A-11

Query Match 4.3%; Score 7; DB 1; Length 132;

Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 31 VTLOGRS 37
|||||||
DB 85 VTLOGRS 91

RESULT 2

US-08-470-298B-11
; Sequence 11, Application US/08470298B

Patent No. 5844081
GENERAL INFORMATION:
APPLICANT: NI, JIAN
APPLICANT: GENTZ, REINER
APPLICANT: YU, GUO-LIANG
APPLICANT: ROSEN, CRAIG A.
TITLE OF INVENTION: CYTOSTATIN I
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUMAN GENOME SCIENCES, INC.
STREET: 9410 KEY WEST AVENUE
CITY: ROCKVILLE
STATE: MD
COUNTRY: US
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,298B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BROOKES, ALLAN A.
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PE175D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8512
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: MYELIN P2 (FIGURE 2)
US-08-470-298B-11
Query Match 4.3%; Score 7; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 31 VTLOGRS 37
|||||||
DB 85 VTLOGRS 91
RESULT 3
US-09-023-073A-11
; Sequence 11, Application US/09023073A
Patent No. 5977309
GENERAL INFORMATION:
APPLICANT: NI, Jian
APPLICANT: GENTZ, REINER
APPLICANT: YU, GUO-LIANG
APPLICANT: ROSEN, CRAIG A.
TITLE OF INVENTION: CYTOSTATIN I
NUMBER OF SEQUENCES: 11

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
;; STREET: 9410 KEY WEST AVENUE
;; CITY: ROCKVILLE
;; STATE: MARYLAND
;; COUNTRY: USA
;; ZIP: 20850
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent in Release #1.0, Version #1.30
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/023.073A
;; FILING DATE: 13-FEB-1998
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wales, Michele M.
;; REGISTRATION NUMBER: P-43,975
;; REFERENCE/DOCKET NUMBER: PF175D2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 301-610-5772
;; TELEFAX: 301-309-8439
;;
;; INFORMATION FOR SEQ ID NO: 11:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 132 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
US-09-023-073A-11

Query Match
Best Local Similarity 4.3%; Score 7; DB 2; Length 132;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 VTLOGRS 37
|111111|
DB 85 VTLOGRS 91

RESULT 4
US-09-361-737-11
; Sequence 11, Application US/09361737
; Patent No. 6287812
; GENERAL INFORMATION:
; APPLICANT: NI, Jian
; APPLICANT: Gentz, Reiner
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Rosen, Craig A
; TITLE OF INVENTION: Cyclostatin I
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/361.737
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/023.073
; FILING DATE: 13-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Wales, Michele M.

;; REGISTRATION NUMBER: P-43,975
;; REFERENCE/DOCKET NUMBER: PF175D2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 301-610-5772
;; TELEFAX: 301-309-8439
;;
;; INFORMATION FOR SEQ ID NO: 11:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 132 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
US-09-361-737-11

Query Match
Best Local Similarity 4.3%; Score 7; DB 4; Length 132;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 VTLOGRS 37
|111111|
DB 85 VTLOGRS 91

RESULT 5
US-08-132-990A-8
; Sequence 8, Application US/08132990A
; Patent No. 5834589
; GENERAL INFORMATION:
; APPLICANT: MERUELO, DANIEL
; APPLICANT: YOSHIMOTO, TAKAYUKI
; TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding Therefor
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/132.990A
; FILING DATE: 07-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/084,729
; FILING DATE: 29-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05569
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/899,075
; FILING DATE: 11-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/806,178
; FILING DATE: 13-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/627,950
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 8105-004-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66441 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-132-990A-8

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 629;
Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 SLIAYLI 105
|||||
Db 501 SLIAYLI 507

RESULT 6
PCT-US92-09382-8
Sequence 8, Application PC/TUS9209382
GENERAL INFORMATION:

APPLICANT: MERUELO, DANIEL
APPLICANT: YOSHIMOTO, TAKAYUKI
TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding
TITLE OF INVENTION: Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/09382
FILING DATE: 19921213

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Livanat, Shmuel
REGISTRATION NUMBER: 33,949
REFERENCE/DOCKET NUMBER: MERUELO=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 628-5197
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 629 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-09382-8

Query Match
Best Local Similarity 100.0%; Score 7; DB 5; Length 629;
Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 SLIAYLI 105
|||||
Db 501 SLIAYLI 507

RESULT 7
US-09-134-001C-4014
Sequence 4014, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4014
LENGTH: 92
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4014

Query Match
Best Local Similarity 100.0%; Score 6; DB 4; Length 92;
Pred. No. 13e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 NLVVDL 13
|||||
Db 6 NLVVDL 11

RESULT 8
US-08-326-117B-13
Sequence 13, Application US/08326117B
Patent No. 5693491
GENERAL INFORMATION:

APPLICANT: BULLA, LEE A.
APPLICANT: JI, TAE
TITLE OF INVENTION: RECEPTOR FOR A BACILLUS THURINGIENSIS
TITLE OF INVENTION: TOXIN
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave. N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/326,117B
FILING DATE: 19-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MILLMAN, ROBERT A.
REGISTRATION NUMBER: 36,217
REFERENCE/DOCKET NUMBER: 7112-0037.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-326-117B-13

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 124;
Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 SAGGLV 94
|||||
Db 15 SAGGLV 20

RESULT 9
US-08-982-129-13

```
Sequence 13, Application US/08982129
Patent No. 6007981
GENERAL INFORMATION:
APPLICANT: BULLA, LEE A.
TITLE OF INVENTION: RECEPTOR FOR A BACILLUS THURINGIENSIS
TITLE OF INVENTION: TOXIN
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FORSTER
STREET: 2000 Pennsylvania Ave. N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/982.129
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/326.117
FILING DATE: 19-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: MILLMAN, ROBERT A
REGISTRATION NUMBER: 36,217
REFERENCE/DOCKET NUMBER: 7112-0037.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-982-129-13

Query Match
Best Local Similarity 3.7%; Score 6; DB 3; Length 124;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 SAGGLV 94
DB 15 SAGGLV 20

RESULT 10
US-08-983-607-51
Sequence 51, Application US/08983607
Patent No. 6140470
GENERAL INFORMATION:
APPLICANT: Xiahong Cai
TITLE OF INVENTION: Human Anti-Tumor Monoclonal Anti-
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Department of Molecular Biophysics
STREET: 266 Whitney Avenue
CITY: New Haven
STATE: Connecticut
COUNTRY: United States of America
ZIP: 06520-8114
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" 1.44 Mb diskette
COMPUTER: IBM PC
```

```
OPERATING SYSTEM: MS DOS
SOFTWARE: Word Processing
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/983.607
FILING DATE: April 27, 1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/IB96/01032
FILING DATE: June 28, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mary M. Krinsky
REGISTRATION NUMBER: 32423
REFERENCE/DOCKET NUMBER: OCR-679
TELECOMMUNICATION INFORMATION:
TELEPHONE: 203-773-9544
TELEFAX: 203-773-1183
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 residues
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: polypeptide
ORIGINAL SOURCE:
ORGANISM: Homo sapiens (melanoma patient
ORGANISM: Immunized with autologous tumor cells)
INDIVIDUAL ISOLATE: peripheral blood lympho-
IMMEDIATE SOURCE: cytes
LIBRARY: VH antibodies obtained from fuses
CLONE: E-13
FEATURE:
NAME/KEY: heavy chain
US-08-983-607-51

Query Match
Best Local Similarity 3.7%; Score 6; DB 4; Length 124;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 SAGGLV 94
DB 7 SAGGLV 12

RESULT 11
US-09-134-001C-4330
Sequence 4330, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucelte-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134.001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4330
LENGTH: 146
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4330

Query Match
Best Local Similarity 3.7%; Score 6; DB 4; Length 146;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 45 NFSQTV 50
 |||||
 Db 77 NFSQTV 82

RESULT 12
 US-08-985-526-36
 ; Sequence 36, Application US/08985526
 ; Patent No. 6080728

GENERAL INFORMATION:
 APPLICANT: Mixson, James A
 TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
 TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
 TITLE OF INVENTION: THERAPY
 NUMBER OF SEQUENCES: 43
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Connolly, Bove, Lodge, & Hutz
 STREET: 1220 Market Street, P.O. Box 2207
 CITY: Wilmington
 STATE: Delaware
 COUNTRY: U.S.A.

ZIP: 19899
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/985,526

FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/608,845
 FILING DATE: 16-JUL-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: McMorris Jr., Robert G
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (302) 658-9141
 TELEFAX: (302) 658-5613
 INFORMATION FOR SEQ ID NO: 36:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 185 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-985-526-36

Query Match 3.7%; Score 6; DB 3; Length 185;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 151 RGSVP1 156
 |||||
 Db 66 RGSVP1 71

RESULT 13
 US-09-561-500-13
 ; Sequence 13, Application US/09561500
 ; Patent No. 6342219

GENERAL INFORMATION:
 APPLICANT: Philip E. Thorpe
 TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
 TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
 FILE REFERENCE: 4001.002500
 CURRENT APPLICATION NUMBER: US/09/561,500
 CURRENT FILING DATE: 2000-04-28
 PRIOR APPLICATION NUMBER: 60/131,432
 PRIOR FILING DATE: 1999-04-28
 NUMBER OF SEQ ID NOS: 44
 SOFTWARE: Patent Ver. 2.0
 SEQ ID NO 13

LENGTH: 191
 TYPE: PRT

ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
 US-09-561-500-13

Query Match 3.7%; Score 6; DB 4; Length 191;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 151 RGSVP1 156
 |||||
 Db 73 RGSVP1 78

RESULT 14
 US-09-561-108-13
 ; Sequence 13, Application US/09561108
 ; Patent No. 6342221

GENERAL INFORMATION:
 APPLICANT: Philip E. Thorpe
 TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VE
 FILE REFERENCE: 4001.002584
 CURRENT APPLICATION NUMBER: US/09/561,108
 CURRENT FILING DATE: 2000-04-28
 PRIOR APPLICATION NUMBER: 60/131,432
 PRIOR FILING DATE: 1999-04-28
 NUMBER OF SEQ ID NOS: 44
 SOFTWARE: Patent Ver. 2.0
 SEQ ID NO 13

LENGTH: 191
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
 US-09-561-108-13

Query Match 3.7%; Score 6; DB 4; Length 191;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 151 RGSVP1 156
 |||||
 Db 73 RGSVP1 78

RESULT 15
 US-09-561-526-13
 ; Sequence 13, Application US/09561526
 ; Patent No. 6418758

GENERAL INFORMATION:
 APPLICANT: Philip E. Thorpe
 TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
 FILE REFERENCE: 4001.002586
 CURRENT APPLICATION NUMBER: US/09/561,526
 CURRENT FILING DATE: 2000-04-28
 PRIOR APPLICATION NUMBER: 60/131,432
 PRIOR FILING DATE: 1999-04-28
 NUMBER OF SEQ ID NOS: 44
 SOFTWARE: Patent Ver. 2.0
 SEQ ID NO 13

LENGTH: 191
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
 US-09-561-526-13

Query Match 3.7%; Score 6; DB 4; Length 191;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 151 RGSVP1 156
DB 73 RGSVP1 78

RESULT 16

US-08-159-784-2
Sequence 2, Application US/08159784
Patent No. 5643783
GENERAL INFORMATION:
APPLICANT: Bjorn R. Olsen
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
City: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION NUMBER: US/08/159,784
APPLICATION NUMBER: 530
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 195
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-159-784-2

Query Match 3.7%: Score 6; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 151 RGSVP1 156
DB 77 RGSVP1 82

RESULT 17

US-08-248-839C-94
Sequence 94, Application US/08248839C
Patent No. 5843702
GENERAL INFORMATION:
APPLICANT: McConnell, David
APPLICANT: Devine, Kevin
APPLICANT: O'Kane, Charles
TITLE OF INVENTION: A Gene Expression System
NUMBER OF SEQUENCES: 185
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58437020 No. 5843702disk of No. 5843702th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA

ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/248,839C
FILING DATE: 25-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valera A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 3614.214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-248-839C-94

Query Match. 3.7%: Score 6; DB 2; Length 213;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 27 ITDYVT 32
DB 148 ITDYVT 153

RESULT 18

US-07-885-089B-4
Sequence 4, Application US/07885089B
Patent No. 5830995
GENERAL INFORMATION:
APPLICANT: Shoyab, Mohammed
APPLICANT: McDonald, Vicki L.
APPLICANT: Bradley, James G.
APPLICANT: Plozman, Gregory D.
TITLE OF INVENTION: AMPHIREGULINS: A FAMILY OF
HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/885,089B
FILING DATE: 18-MAY-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 5624-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

LENGTH: 243 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-885-089B-4

Query Match 3.7%; Score 6; DB 2; Length 243;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 89 SAGSLV 94
|||||
Db 45 SAGSLV 50

RESULT 19
US-07-885-089B-9
; Sequence 9, Application US/07885089B
; Patent No. 5830995
; GENERAL INFORMATION:
; APPLICANT: Shoyab, Mohammed
; APPLICANT: McDonald, Vicki L.
; APPLICANT: Bradley, James G.
; APPLICANT: Plowman, Gregory D.
; TITLE OF INVENTION: AMPHIREGULINS: A FAMILY OF
; TITLE OF INVENTION: HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,089B
; FILING DATE: 18-MAY-1992
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 5624-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-07-885-089B-9

Query Match 3.7%; Score 6; DB 2; Length 243;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 89 SAGSLV 94
|||||
Db 45 SAGSLV 50

RESULT 20
US-08-944-483-70
; Sequence 70, Application US/08944483
; Patent No. 6232456

GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROBE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
US-08-944-483-70

Query Match 3.7%; Score 6; DB 4; Length 243;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 97 AGSLIA 102
|||||
Db 26 AGSLIA 31

RESULT 21
5202428-2
; Patent No. 5202428
; APPLICANT: SCHUBERT, DAVID
; TITLE OF INVENTION: DNA ENCODING NEUROTROPHIC GROWTH FACTOR
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/590,359
; FILING DATE: 27-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 541,276
; FILING DATE: 20-JUN-1990
; SEQ ID NO: 2;
; LENGTH: 243
5202428-2

Query Match 3.7%; Score 6; DB 6; Length 243;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 SAGGLV 94
111111
DB 45 SAGGLV 50

RESULT 22

US-08-956-267A-2
Sequence 2, Application US/08956267A
Patent No. 5945328
GENERAL INFORMATION:
APPLICANT: MOLDRE, Helle Fabricius
APPLICANT: KJELDSEN, Thomas Borglum
TITLE OF INVENTION: A Process for Producing Trypsin
TITLE OF INVENTION: (trypsinogen)
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 5945328 of No. 5945328 of No. 5945328 of America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,267A
FILING DATE: 22-OCT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROZEK, Carol
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 4500.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 247 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-956-267A-2

Query Match 3.7%; Score 6; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSSSY 58
111111
DB 151 SSSSY 156

RESULT 23

US-08-989-386-5
Sequence 5, Application US/08989386
Patent No. 5989860
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hillman, Jennifer L.
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: HUMAN ISOMERASE HOMOLOGS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/989,386
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0443 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 249 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: KIDNOT09
CLONE: 1419071
US-08-989-386-5

Query Match 3.7%; Score 6; DB 2; Length 249;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 DVSARD 142
111111
DB 125 DVSARD 130

RESULT 24

US-09-238-303-10
Sequence 10, Application US/09238303B
Patent No. 6284253
GENERAL INFORMATION:
APPLICANT: Barr, Margaret C.
TITLE OF INVENTION: No. 6284253el feline Immunodeficiency Virus Nucleotide Sequence
FILE REFERENCE: 18617.0059
CURRENT APPLICATION NUMBER: US/09/238,303B
CURRENT FILING DATE: 1999-01-28
EARLIER APPLICATION NUMBER: US 60/072,927
EARLIER FILING DATE: 1998-01-29
NUMBER OF SEQ ID NOS: 17
SEQ ID NO 10
LENGTH: 252
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: protein encoded by the vif gene of a recombinant viral
OTHER INFORMATION: clone constructed from the genomic DNA of a Pallas's cat feline
US-09-238-303-10

Query Match 3.7%; Score 6; DB 4; Length 252;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 GLVTKA 97
|||||
DB 154 GLVTKA 159

RESULT 25

US-08-491-944-2
; Sequence 2, Application US/08491944
; Patent No. 6361982

GENERAL INFORMATION:

APPLICANT: Walters, Nicola Jane
Barton, Barry
Earl, Alison J.

TITLE OF INVENTION: No. 6361982el Compounds

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESSES:

ADDRESSEE: SmithKline Beecham P.L.C.

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/491,944

FILING DATE: 03-Aug-1995

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Gimmil, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P31457-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 262 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Streptomyces clavuligerus

STRAIN: S. clavuligerus 27064

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Query Match 3.7%; Score 6; DB 4; Length 262;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 GSVPIP 157
|||||
DB 20 GSVPIP 25

RESULT 26

US-09-025-769B-178

; Sequence 178, Application US/09025769B

; Patent No. 6300064

GENERAL INFORMATION:

APPLICANT: Knappik, Achim

APPLICANT: Pack, Peter

APPLICANT: Ilag, Vic

APPLICANT: Ge, Liming

APPLICANT: Moroney, Simon

APPLICANT: Plueckthun, Andreas

TITLE OF INVENTION: Protein/(Poly)peptide libraries

NUMBER OF SEQUENCES: 373

CORRESPONDENCE ADDRESS:

ADDRESSEE: James F. Haley, Jr., Esq. c/o Fish & Neave

STREET: 1251 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10021

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/025,769B

FILING DATE: 18-FEB-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 95 11 3021.0

FILING DATE: 18-AUG-1995

ATTORNEY/AGENT INFORMATION:

NAME: James F. Haley, Jr., Esq.

REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: MORPHO/5

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)596-9000

TELEFAX: (212)596-9090

INFORMATION FOR SEQ ID NO: 178:

SEQUENCE CHARACTERISTICS:

LENGTH: 281 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

Query Match 3.7%; Score 6; DB 4; Length 281;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 VSSAG 92
|||||
DB 143 VSSAG 148

RESULT 27

US-09-512-342-14

; Sequence 14, Application US/09512342

; Patent No. 638068

GENERAL INFORMATION:

APPLICANT: SATOH, SHINOBU

APPLICANT: MASUDA, SUSUMU

TITLE OF INVENTION: METHOD FOR PRODUCING FOREIGN POLYPEPTIDE IN PLANT

FILE REFERENCE: 081356/0142

CURRENT APPLICATION NUMBER: US/09/512,342

CURRENT FILING DATE: 2000-02-24

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 14

LENGTH: 293

TYPE: PRT

ORGANISM: Cucumis sativus

US-09-512-342-14

Query Match 3.7%; Score 6; DB 4; Length 293;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 DVTI 34
|||||
DB 144 DVTI 149

RESULT 28

US-08-118-270-57
; Sequence 57, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
; APPLICANT: Murphy, Randall B.
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,270
; FILING DATE: 09-SEP-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-118-270-57

Query Match 3.7%; Score 6; DB 1; Length 297;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSSYP 58
DB 150 SSSYP 155

RESULT 29
US-08-118-270-58
; Sequence 58, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
; APPLICANT: Murphy, Randall B.
; APPLICANT: Schuster, David I.
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,270
; FILING DATE: 09-SEP-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-118-270-57

OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,270
; FILING DATE: 09-SEP-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-118-270-58

Query Match 3.7%; Score 6; DB 1; Length 297;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSSYP 58
DB 149 SSSYP 154

RESULT 30
PCT-US93-08528-57
; Sequence 57, Application PC/TUS9308528
; GENERAL INFORMATION:
; APPLICANT: New York University
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08528
; FILING DATE: 09-SEP-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-08528-57

Query Match
Best Local Similarity 3.7%; Score 6; DB 5; Length 297;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 SSSYP 58
DB 150 SSSYP 155

RESULT 31
PCT-US93-08528-58
Sequence 58, Application PC/TUS9308528
GENERAL INFORMATION:
APPLICANT: New York University
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08528
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY=2 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-08528-58

Query Match
Best Local Similarity 3.7%; Score 6; DB 5; Length 297;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 SSSYP 58
DB 149 SSSYP 154

RESULT 32
US-09-134-001C-3238
Sequence 3238, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3238
LENGTH: 307
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3238

Query Match
Best Local Similarity 3.7%; Score 6; DB 4; Length 307;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 YNSDP 117
DB 192 YNSDP 197

RESULT 33
US-09-424-349A-8
Sequence 8, Application US/09424349A
Patent No. 6387668
GENERAL INFORMATION:
APPLICANT: Lutz Spelberg, Jeffrey Harald
Rink, Rick
Kellogg, Richard Morrison
Janssen, Dick Barend
TITLE OF INVENTION: Enantioselective epoxide hydrolases and genes
encoding these.
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Vereenigde
STREET: Nieuwe Parklaan 97
CITY: The Hague
STATE: Zuid-Holland
COUNTRY: The Netherlands
ZIP: 2587 BN
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/424,349A
FILING DATE: 23-Feb-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 97201515.0
FILING DATE: 21-MAY-1997
APPLICATION NUMBER: PCT/NL98/00290
FILING DATE: 20-MAY-1998
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-424-349A-8

Query Match
Best Local Similarity 3.7%; Score 6; DB 4; Length 329;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 VLSNFS 47
DB 127 VLSNFS 132

RESULT 34
US-08-164-292B-13
; Sequence 13, Application US/08164292B
; Patent No. 5820868
; GENERAL INFORMATION:
; APPLICANT: MITTAL, SURESH K.
; APPLICANT: GRAHAM, FRANK L.
; APPLICANT: PREVEC, LUDVIG
; APPLICANT: BABIUK, LORNE A.
; TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE
; TITLE OF INVENTION: ADENOVIRUS EXPRESSION VECTOR SYSTEM
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 345 California Street
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104-2675
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/164,292B
; FILING DATE: 09-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GRACEY, NANCY J.
; REGISTRATION NUMBER: 28,216
; REFERENCE/DOCKET NUMBER: 29310-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 677-7000
; TELEFAX: (415) 677-7522
; TELEX: 34-0154
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 347 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-164-292B-13

Query Match 3.7%; Score 6; DB 2; Length 347;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 NFGTV 50
DB 96 NFGTV 101

RESULT 35
US-08-845-623-13
; Sequence 13, Application US/08845623A
; Patent No. 6001591
; GENERAL INFORMATION:
; APPLICANT: BABIUK, LORNE A.
; APPLICANT: TIKOO, SURESH K.
; APPLICANT: REDDY, POLICE S.
; TITLE OF INVENTION: BOVINE ADENOVIRUS 3 GENOME
; FILE REFERENCE: 293102002120
; CURRENT APPLICATION NUMBER: US/08/845,623A
; CURRENT FILING DATE: 1997-04-25
; EARLIER APPLICATION NUMBER: 08/164,294
; EARLIER FILING DATE: 1993-12-09
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13

; LENGTH: 347
; TYPE: PRT
; ORGANISM: Human adenovirus type 5
US-08-845-623-13

Query Match 3.7%; Score 6; DB 3; Length 347;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 NFGTV 50
DB 96 NFGTV 101

RESULT 36
US-08-815-927-13
; Sequence 13, Application US/08815927
; Patent No. 6086890
; GENERAL INFORMATION:
; APPLICANT: MITTAL, SURESH K.
; APPLICANT: GRAHAM, FRANK L.
; APPLICANT: PREVEC, LUDVIG
; APPLICANT: BABIUK, LORNE A.
; TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE ADENOVIRUS EXPRESSION
; FILE REFERENCE: 293102002101
; CURRENT APPLICATION NUMBER: US/08/815,927
; CURRENT FILING DATE: 1997-03-13
; EARLIER APPLICATION NUMBER: 08/164,294
; EARLIER FILING DATE: 1993-12-09
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Human adenovirus type 5
US-08-815-927-13

Query Match 3.7%; Score 6; DB 3; Length 347;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 NFGTV 50
DB 96 NFGTV 101

RESULT 37
US-09-103-330-13
; Sequence 13, Application US/09103330A
; Patent No. 6319716
; GENERAL INFORMATION:
; APPLICANT: TIKOO, SURESH K.
; APPLICANT: BABIUK, LORNE A.
; APPLICANT: REDDY, POLICE S.
; TITLE OF INVENTION: ISOLATION OF MUTANTS IN THE E3 REGION OF THE
; TITLE OF INVENTION: BOVINE ADENOVIRUS GENOME AND THEIR USE IN VACCINES
; FILE REFERENCE: 293102002121
; CURRENT APPLICATION NUMBER: US/09/103,330A
; CURRENT FILING DATE: 1998-06-23
; EARLIER APPLICATION NUMBER: 08/880,234
; EARLIER FILING DATE: 1997-06-23
; EARLIER APPLICATION NUMBER: 08/164,292
; EARLIER FILING DATE: 1993-12-09
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Human adenovirus type 5
US-09-103-330-13

Query Match 3.7%; Score 6; DB 4; Length 347;

Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 NFSGTV 50
Db 96 NFSGTV 101

RESULT 38
US-09-435-242-13
; Sequence 13, Application US/09435242
; Patent No. 6379944
; GENERAL INFORMATION:
; APPLICANT: MITTAL, SURESH K.
; APPLICANT: GRAHAM, FRANK L.
; APPLICANT: PREVIC, LUDVIG
; APPLICANT: BABURK, LORNE A.
; TITLE OF INVENTION: MAMMALIAN CELL LINES EXPRESSING BOVINE ADENOVIRUS FUNCTIONS
; FILE REFERENCE: 29310202102
; CURRENT APPLICATION NUMBER: US/09/435,242
; CURRENT FILING DATE: 1999-11-05
; EARLIER APPLICATION NUMBER: 08/815,927
; EARLIER FILING DATE: 1997-03-13
; EARLIER APPLICATION NUMBER: 08/164,294
; EARLIER FILING DATE: 1993-12-09
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Human adenovirus type 5
US-09-435-242-13

Query Match 3.7%; Score 6; DB 4; Length 347;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 NFSGTV 50
Db 96 NFSGTV 101

RESULT 39
US-09-615-192A-338
; Sequence 338, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003c4U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 338
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-615-192A-338

Query Match 3.7%; Score 6; DB 4; Length 358;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GSSYPF 59

Db 45 GSSYPF 50

RESULT 40
US-09-134-001C-4365
; Sequence 4365, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4365
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4365

Query Match 3.7%; Score 6; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 YLTPVS 88
Db 39 YLTPVS 44

RESULT 41
US-09-134-001C-4259
; Sequence 4259, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4259
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4259

Query Match 3.7%; Score 6; DB 4; Length 376;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 NDYPET 26
Db 247 NDYPET 252

RESULT 42
US-08-689-421-23
; Sequence 23, Application US/08689421
; Patent No. 6008029
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie S.
; APPLICANT: Brown, Kimberley M.

APPLICANT: Kaupinen, Sakari
APPLICANT: Halkier, Torben P
TITLE OF INVENTION: PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS
TITLE OF INVENTION: ENCODING SAME
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESS: No. 60080290 No. 6008029disk of No. 6008029th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/689,421
FILING DATE: 9-AUG-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4554.204-WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 387 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-689-421-23

Query Match
Best Local Similarity 3.7%; Score 6; DB 3; Length 387;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VVWVGQ 7
Db 361 VVWVGQ 366

RESULT 43
US-09-389-528-23
Sequence 23, Application US/09389528
Patent No. 6207430
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie S.
APPLICANT: Brown, Kimberly M.
APPLICANT: Kaupinen, Sakari
APPLICANT: Halkier, Torben P
TITLE OF INVENTION: PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS
TITLE OF INVENTION: ENCODING SAME
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESS: No. 62074300 No. 6207430disk of No. 6207430th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/389,528
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/689,421
FILING DATE: 9-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4554.204-WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 387 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-389-528-23

Query Match
Best Local Similarity 3.7%; Score 6; DB 4; Length 387;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VVWVGQ 7
Db 361 VVWVGQ 366

RESULT 44
US-09-181-827A-23
Sequence 23, Application US/09181827A
Patent No. 6242232
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie S.
APPLICANT: Brown, Kimberly M.
APPLICANT: Kaupinen, Sakari
APPLICANT: Halkier, Torben P
TITLE OF INVENTION: Purified Coprinus Laccases And Nucleic
Acids Encoding Same
FILE REFERENCE: 4554.200-US
CURRENT APPLICATION NUMBER: US/09/181,827A
CURRENT FILING DATE: 1998-10-28
PRIOR APPLICATION NUMBER: 60/002,800
PRIOR FILING DATE: 1995-08-25
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 387
TYPE: PRT
ORGANISM: Coprinus cinereus
US-09-181-827A-23

Query Match
Best Local Similarity 3.7%; Score 6; DB 4; Length 387;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VVWVGQ 7
Db 361 VVWVGQ 366

RESULT 45
US-09-131-028A-4
Sequence 4, Application US/09131028A
Patent No. 6287866
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Mukerji, Pradip
APPLICANT: Lemmel, Steven A.
APPLICANT: Leonard, Amanda Eun-Yeong
APPLICANT: Chaudhary, Sunita
TITLE OF INVENTION: BETA-CASEIN EXPRESSING CONSTRUCTS
FILE REFERENCE: 6004.US.PI
CURRENT APPLICATION NUMBER: US/09/131,028A

; CURRENT FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: US 08/064,440
; PRIOR FILING DATE: 1993-05-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-131-028A-4

Query Match 3.7%; Score 6; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 TPVSSA 90
|||||
Db 344 TPVSSA 349

Search completed: November 28, 2002, 19:06:41
Job time : 25 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 19:05:41 ; Search time 13 Seconds

(without alignments)

197.215 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186

Perfect score: 161

Sequence: 1 PVTNNGQNLVVDLSTOIFCH.....DVTVLPDYGSRVPLTVY 161

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 102317 seqs, 15924203 residues

Word size : 0

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	161	100.0	279	10	US-09-900-575-29
2	92	57.1	279	10	US-09-900-575-34
3	92	57.1	279	10	US-09-900-575-36
4	92	57.1	279	10	US-09-900-575-44
5	92	57.1	300	10	US-09-912-020-367
6	90	55.9	279	10	US-09-900-575-27
7	90	55.9	279	10	US-09-900-575-28
8	90	55.9	279	10	US-09-900-575-32
9	90	55.9	279	10	US-09-900-575-39
10	90	55.9	279	10	US-09-900-575-43
11	90	55.9	279	10	US-09-900-575-45
12	75	46.6	280	10	US-09-900-575-30
13	67	41.6	279	10	US-09-900-575-25
14	67	41.6	279	10	US-09-900-575-26
15	67	41.6	279	10	US-09-900-575-35
16	67	41.6	279	10	US-09-900-575-37
17	67	41.6	279	10	US-09-900-575-40
18	67	41.6	279	10	US-09-900-575-41
19	67	41.6	279	10	US-09-900-575-42

20	67	41.6	279	10	US-09-900-575-45	Sequence 45, App1
21	60	37.3	279	10	US-09-900-575-38	Sequence 38, App1
22	56	34.8	279	10	US-09-900-575-33	Sequence 33, App1
23	48	29.8	279	10	US-09-900-575-23	Sequence 23, App1
24	43	26.7	279	10	US-09-900-575-24	Sequence 24, App1
25	43	26.7	279	10	US-09-900-575-31	Sequence 31, App1
26	8	5.0	1036	12	US-10-014-882-2	Sequence 2, App1
27	7	4.3	132	10	US-09-901-436A-11	Sequence 11, App1
28	7	4.3	1040	10	US-09-912-020-282	Sequence 282, App
29	7	4.3	1040	10	US-09-815-242-10203	Sequence 10203, A
30	6	3.7	40	10	US-09-864-761-37358	Sequence 37358, A
31	6	3.7	46	10	US-09-726-643-65	Sequence 65, App1
32	6	3.7	54	10	US-09-864-761-34310	Sequence 34310, A
33	6	3.7	62	10	US-09-864-761-35614	Sequence 35614, A
34	6	3.7	93	10	US-09-867-550-1142	Sequence 1142, Ap
35	6	3.7	93	10	US-09-867-550-1428	Sequence 1428, A
36	6	3.7	112	10	US-09-864-761-43168	Sequence 43168, A
37	6	3.7	144	10	US-09-728-914-10	Sequence 10, App1
38	6	3.7	183	10	US-09-739-907-65	Sequence 65, App1
39	6	3.7	185	12	US-10-036-869-36	Sequence 36, App1
40	6	3.7	192	10	US-09-998-831-13	Sequence 13, App1
41	6	3.7	192	10	US-09-925-302-728	Sequence 728, App
42	6	3.7	194	10	US-09-734-017A-74	Sequence 74, App1
43	6	3.7	213	10	US-09-815-242-13993	Sequence 13993, A
44	6	3.7	252	10	US-09-946-239-10	Sequence 10, App1
45	6	3.7	275	9	US-09-828-523A-4	Sequence 4, App1
46	6	3.7	284	10	US-09-828-523A-54	Sequence 54, App1
47	6	3.7	305	10	US-09-443-704-8	Sequence 8, App1
48	6	3.7	335	10	US-09-925-300-1575	Sequence 1575, Ap
49	6	3.7	352	10	US-09-771-956-23	Sequence 23, App1
50	6	3.7	353	10	US-09-962-739-4	Sequence 4, App1
51	6	3.7	364	10	US-09-922-494-10	Sequence 10, App1
52	6	3.7	364	10	US-09-922-494-11	Sequence 11, App1
53	6	3.7	364	10	US-09-925-100-1391	Sequence 1391, Ap
54	6	3.7	365	10	US-09-784-358-6	Sequence 6, App1
55	6	3.7	372	10	US-09-801-368-206	Sequence 206, App
56	6	3.7	376	10	US-09-771-956-25	Sequence 25, App1
57	6	3.7	395	10	US-09-835-270-4	Sequence 4, App1
58	6	3.7	431	10	US-09-784-358-4	Sequence 4, App1
59	6	3.7	446	10	US-09-930-521-7	Sequence 7, App1
60	6	3.7	448	10	US-09-930-521-6	Sequence 6, App1
61	6	3.7	453	12	US-10-083-452-6	Sequence 2, App1
62	6	3.7	455	10	US-09-962-646-2	Sequence 2, App1
63	6	3.7	460	9	US-08-813-453A-39	Sequence 39, App1
64	6	3.7	466	10	US-09-765-205-28	Sequence 28, App1
65	6	3.7	499	10	US-09-771-956-24	Sequence 24, App1
66	6	3.7	508	10	US-09-771-956-22	Sequence 22, App1
67	6	3.7	513	10	US-09-833-745-55	Sequence 55, App1
68	6	3.7	514	9	US-09-712-363-154	Sequence 154, App
69	6	3.7	518	10	US-09-803-589-10	Sequence 10, App1
70	6	3.7	523	10	US-09-930-521-4	Sequence 4, App1
71	6	3.7	541	10	US-09-981-421-4	Sequence 4, App1
72	6	3.7	541	12	US-10-157-447-2	Sequence 2, App1
73	6	3.7	551	10	US-10-105-929-16	Sequence 16, App1
74	6	3.7	576	10	US-09-851-686-6	Sequence 6, App1
75	6	3.7	592	9	US-09-813-453A-22	Sequence 22, App1
76	6	3.7	592	9	US-09-813-453A-43	Sequence 43, App1
77	6	3.7	605	9	US-09-906-209-4	Sequence 25, App1
78	6	3.7	612	12	US-10-001-851-25	Sequence 10, App1
79	6	3.7	650	10	US-09-784-358-10	Sequence 10, App1
80	6	3.7	655	9	US-09-906-209-10	Sequence 2, App1
81	6	3.7	679	10	US-09-962-739-2	Sequence 2, App1
82	6	3.7	679	10	US-09-925-302-632	Sequence 632, App
83	6	3.7	724	10	US-09-784-358-8	Sequence 8, App1
84	6	3.7	727	10	US-09-443-023A-12	Sequence 12, App1
85	6	3.7	765	10	US-09-975-326-4	Sequence 4, App1
86	6	3.7	766	10	US-09-975-326-4	Sequence 2, App1
87	6	3.7	771	10	US-09-784-358-14	Sequence 14, App1
88	6	3.7	802	10	US-09-838-561-2	Sequence 2, App1
89	6	3.7	802	10	US-09-816-760-2	Sequence 12, App1
90	6	3.7	845	10	US-09-784-358-12	Sequence 4, App1
91	6	3.7	950	10	US-09-321-967B-4	Sequence 3, App1
92	6	3.7	969	8	US-08-945-749-3	Sequence 3, App1


```
969 5 3.1 481 12 US-10-105-929-8 Sequence 8, Appli
970 5 3.1 482 10 US-09-779-307-17 Sequence 17, Appl
971 5 3.1 482 10 US-09-861-400-5 Sequence 5, Appl
972 5 3.1 483 9 US-09-974-298-41 Sequence 41, Appl
973 5 3.1 483 9 US-09-981-353-158 Sequence 158, App
974 5 3.1 483 10 US-09-739-254-154 Sequence 154, App
975 5 3.1 483 10 US-09-732-618-10 Sequence 10, Appl
976 5 3.1 483 10 US-09-904-615-154 Sequence 154, App
977 5 3.1 483 10 US-09-805-458A-6 Sequence 6, Appl
978 5 3.1 483 10 US-09-919-497-79 Sequence 79, Appl
979 5 3.1 483 12 US-10-042-991-4 Sequence 4, Appl
980 5 3.1 484 10 US-09-790-264-36 Sequence 36, Appl
981 5 3.1 484 10 US-09-815-242-11054 Sequence 11054, A
982 5 3.1 484 12 US-10-006-867-78 Sequence 78, Appl
983 5 3.1 484 12 US-10-052-586-280 Sequence 280, App
984 5 3.1 485 9 US-09-861-846-10 Sequence 10, Appl
985 5 3.1 485 10 US-09-861-289-10 Sequence 10, Appl
986 5 3.1 487 10 US-09-815-242-5183 Sequence 5183, Ap
987 5 3.1 487 10 US-09-825-414-66 Sequence 66, Appl
988 5 3.1 488 10 US-09-884-260A-38 Sequence 38, Appl
989 5 3.1 488 12 US-10-042-991-6 Sequence 6, Appl
990 5 3.1 491 10 US-09-881-752A-128 Sequence 128, App
991 5 3.1 493 10 US-09-815-242-12052 Sequence 12052, A
992 5 3.1 494 10 US-09-815-242-10507 Sequence 10507, A
993 5 3.1 494 12 US-10-143-002-4 Sequence 4, Appl
994 5 3.1 496 9 US-09-963-234-2 Sequence 2, Appl
995 5 3.1 496 10 US-09-974-592-10 Sequence 10, Appl
996 5 3.1 497 10 US-09-734-676-2 Sequence 2, Appl
997 5 3.1 498 10 US-09-925-302-475 Sequence 475, App
998 5 3.1 498 10 US-09-764-864-1122 Sequence 1122, Ap
999 5 3.1 499 10 US-09-734-676-4 Sequence 4, Appl
1000 5 3.1 500 9 US-10-029-009-19 Sequence 19, Appl
```

ALIGNMENTS

```
RESULT 1
US-09-900-575-29
; Sequence 29, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 29
; LENGTH: 279
; TYPE: PR
; ORGANISM: E. coli
US-09-900-575-29

Query Match          100.0%; Score 161; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 4.3e-148;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 PVTNNGQNLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLNFGTGYKSGSSYFP 60
Db 26 PVTNNGQNLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLNFGTGYKSGSSYFP 85
QY 61 TTSETPRVYNSRTDKPWPVALYLPVSSAGLVIKAGSLIAVLILRQTNVNSDDFOFV 120
Db 86 TTSETPRVYNSRTDKPWPVALYLPVSSAGLVIKAGSLIAVLILRQTNVNSDDFOFV 145
QY 121 WNIYANDVYVPTGGCDVSAARDVYTLTPDYRGSVPIPLTVY 161
```

```
Db 146 WNIYANDVYVPTGGCDVSAARDVYTLTPDYRGSVPIPLTVY 186
|||||
RESULT 2
US-09-900-575-34
; Sequence 34, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 279
; TYPE: PR
; ORGANISM: E. coli
US-09-900-575-34

Query Match          57.1%; Score 92; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.7e-81;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 PVTNNGQNLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLNFGTGYKSGSSYFP 60
Db 26 PVTNNGQNLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLNFGTGYKSGSSYFP 85
QY 61 TTSETPRVYNSRTDKPWPVALYLPVSSAG 92
Db 86 TTSETPRVYNSRTDKPWPVALYLPVSSAG 117
QY 121 WNIYANDVYVPTGGCDVSAARDVYTLTPDYRGSVPIPLTVY 161
Db 146 WNIYANDVYVPTGGCDVSAARDVYTLTPDYRGSVPIPLTVY 186
|||||
RESULT 3
US-09-900-575-36
; Sequence 36, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 36
; LENGTH: 279
; TYPE: PR
; ORGANISM: E. coli
US-09-900-575-36

Query Match          57.1%; Score 92; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.7e-81;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 PVTNNGQNLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLNFGTGYKSGSSYFP 60
Db 26 PVTNNGQNLVVDLSTQIFCHNDYPETITDYVTLQSGAYGVLNFGTGYKSGSSYFP 85
QY 61 TTSETPRVYNSRTDKPWPVALYLPVSSAG 92
Db 86 TTSETPRVYNSRTDKPWPVALYLPVSSAG 117
QY 121 WNIYANDVYVPTGGCDVSAARDVYTLTPDYRGSVPIPLTVY 161
Db 146 WNIYANDVYVPTGGCDVSAARDVYTLTPDYRGSVPIPLTVY 186
|||||
```

Db 86 TTSETPRVYNSRTDKPMPVALYLTVPSSAGG 117

RESULT 4

US-09-900-575-44
; Sequence 44, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burteln, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-44

Query Match 57.1%; Score 92; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.7e-81;

Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVVWGNLVVDLSTOIFCHNDYPERITDVTYLOKRSAYGVLSNFGTVKYGSSYPPT 60

Db 26 PVVWGNLVVDLSTOIFCHNDYPERITDVTYLOKRSAYGVLSNFGTVKYGSSYPPT 85

OY 61 TTSETPRVYNSRTDKPMPVALYLTVPSSAGG 92

Db 86 TTSETPRVYNSRTDKPMPVALYLTVPSSAGG 117

RESULT 5

US-09-912-020-367
; Sequence 367, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DIV1
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 367
; LENGTH: 300
; TYPE: PRT
; ORGANISM: E. coli
US-09-912-020-367

Query Match 57.1%; Score 92; DB 10; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.8e-81;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVVWGNLVVDLSTOIFCHNDYPERITDVTYLOKRSAYGVLSNFGTVKYGSSYPPT 60
Db 47 PVVWGNLVVDLSTOIFCHNDYPERITDVTYLOKRSAYGVLSNFGTVKYGSSYPPT 106
OY 61 TTSETPRVYNSRTDKPMPVALYLTVPSSAGG 92
Db 107 TTSETPRVYNSRTDKPMPVALYLTVPSSAGG 138

RESULT 6

US-09-900-575-27
; Sequence 27, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burteln, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-27

Query Match 55.9%; Score 90; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.4e-79;

Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VVWGNLVVDLSTOIFCHNDYPERITDVTYLOKRSAYGVLSNFGTVKYGSSYPPT 62

Db 28 VVWGNLVVDLSTOIFCHNDYPERITDVTYLOKRSAYGVLSNFGTVKYGSSYPPT 87

OY 63 SETPRVYNSRTDKPMPVALYLTVPSSAGG 92

Db 88 SETPRVYNSRTDKPMPVALYLTVPSSAGG 117

RESULT 7

US-09-900-575-28
; Sequence 28, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burteln, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-28

Query Match 55.9%; Score 90; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.4e-79;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VVWGNLVVDLSTOIFCHNDYPERITDVTYLOKRSAYGVLSNFGTVKYGSSYPPT 62

Db 28 VNVGQNLVVDLSTQIFCHNDYPETITDVTYLQSGAYGVLNFSGTVKXSGSSYPFPTT 87

QY 63 SETPRVYNSRTDKPMPVALYLTTPVSSAGG 92
88 SETPRVYNSRTDKPMPVALYLTTPVSSAGG 117

RESULT 8

US-09-900-575-32
; Sequence 32, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 32
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-32

Query Match 55.9%; Score 90; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.4e-79;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VNVGQNLVVDLSTQIFCHNDYPETITDVTYLQSGAYGVLNFSGTVKXSGSSYPFPTT 62
Db 28 VNVGQNLVVDLSTQIFCHNDYPETITDVTYLQSGAYGVLNFSGTVKXSGSSYPFPTT 87
QY 63 SETPRVYNSRTDKPMPVALYLTTPVSSAGG 92
88 SETPRVYNSRTDKPMPVALYLTTPVSSAGG 117

RESULT 9

US-09-900-575-39
; Sequence 39, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 39
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-39

Query Match 55.9%; Score 90; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.4e-79;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 VNVGQNLVVDLSTQIFCHNDYPETITDVTYLQSGAYGVLNFSGTVKXSGSSYPFPTT 62

Db 28 VNVGQNLVVDLSTQIFCHNDYPETITDVTYLQSGAYGVLNFSGTVKXSGSSYPFPTT 87

QY 63 SETPRVYNSRTDKPMPVALYLTTPVSSAGG 92
88 SETPRVYNSRTDKPMPVALYLTTPVSSAGG 117

RESULT 10

US-09-900-575-43
; Sequence 43, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 43
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-43

Query Match 55.9%; Score 90; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.4e-79;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VNVGQNLVVDLSTQIFCHNDYPETITDVTYLQSGAYGVLNFSGTVKXSGSSYPFPTT 62
Db 28 VNVGQNLVVDLSTQIFCHNDYPETITDVTYLQSGAYGVLNFSGTVKXSGSSYPFPTT 87
QY 63 SETPRVYNSRTDKPMPVALYLTTPVSSAGG 92
88 SETPRVYNSRTDKPMPVALYLTTPVSSAGG 117

RESULT 11

US-09-900-575-55
; Sequence 55, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 55
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence of Film proteins for SEQ ID NO: 23 to 45
US-09-900-575-55

Query Match 55.9%; Score 90; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.4e-79;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 VNVGQNLVVDLSTQIFCHNDYPETITDVTYLQSGAYGVLNFSGTVKXSGSSYPFPTT 62

Db 28 VVWGQNLVVDSTQIFCHNDYPERITDVTLQGSAGYGLSNFSGYTKYSGSSYPPTT 87
OY 63 SETPRVYNSRDKPWPVALYLTTPVSSAG 92
Db 88 SETPRVYNSRDKPWPVALYLTTPVSSAG 117

RESULT 12

US-09-900-575-30
; Sequence 30, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burtelin, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 280
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-30

Query Match 46.6%; Score 75; DB 10; Length 280;
Best Local Similarity 100.0%; Pred. No. 4.3e-65;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 FCHNDYPERITDVTLQGSAGYGLSNFSGYTKYSGSSYPPTTSETPRVYNSRDKP 77
Db 43 FCHNDYPERITDVTLQGSAGYGLSNFSGYTKYSGSSYPPTTSETPRVYNSRDKP 102
OY 78 WPVALYLTTPVSSAG 92
Db 103 WPVALYLTTPVSSAG 117

RESULT 13

US-09-900-575-25
; Sequence 25, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burtelin, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-25

Query Match 41.6%; Score 67; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 2.3e-57;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 IKAGSLIIVLLRQTNNNNSDDQFVWNIYANDVYVPTGGCDVSARDVYVTLDPYRGSV 154
Db 120 IKAGSLIIVLLRQTNNNNSDDQFVWNIYANDVYVPTGGCDVSARDVYVTLDPYRGSV 179

Db 120 IKAGSLIIVLLRQTNNNNSDDQFVWNIYANDVYVPTGGCDVSARDVYVTLDPYRGSV 179
OY 155 PIPLTVY 161
Db 180 PIPLTVY 186

RESULT 14

US-09-900-575-26
; Sequence 26, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burtelin, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-26

Query Match 41.6%; Score 67; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 2.3e-57;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 IKAGSLIIVLLRQTNNNNSDDQFVWNIYANDVYVPTGGCDVSARDVYVTLDPYRGSV 154
Db 120 IKAGSLIIVLLRQTNNNNSDDQFVWNIYANDVYVPTGGCDVSARDVYVTLDPYRGSV 179
OY 155 PIPLTVY 161
Db 180 PIPLTVY 186

RESULT 15

US-09-900-575-35
; Sequence 35, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burtelin, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 35
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-35

Query Match 41.6%; Score 67; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 2.3e-57;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 IKAGSLIIVLLRQTNNNNSDDQFVWNIYANDVYVPTGGCDVSARDVYVTLDPYRGSV 154
Db 120 IKAGSLIIVLLRQTNNNNSDDQFVWNIYANDVYVPTGGCDVSARDVYVTLDPYRGSV 179

OY 155 PIPPLTVY 161
| | | | |
Db 180 PIPPLTVY 186

RESULT 16
US-09-900-575-37
; Sequence 37, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burtelin, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 37
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-37

Query Match 41.6%; Score 67; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 2.3e-57;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 IKAGSLI AVLILRQTNNYNSDDFQFWMNIYANDVVPPTGCGDV SAR DVTYTL PDYRGSV 154
| | | | |
Db 120 IKAGSLI AVLILRQTNNYNSDDFQFWMNIYANDVVPPTGCGDV SAR DVTYTL PDYRGSV 179

OY 155 PIPPLTVY 161
| | | | |
Db 180 PIPPLTVY 186

RESULT 17
US-09-900-575-40
; Sequence 40, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burtelin, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 40
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-40

Query Match 41.6%; Score 67; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 2.3e-57;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 IKAGSLI AVLILRQTNNYNSDDFQFWMNIYANDVVPPTGCGDV SAR DVTYTL PDYRGSV 154
| | | | |
Db 120 IKAGSLI AVLILRQTNNYNSDDFQFWMNIYANDVVPPTGCGDV SAR DVTYTL PDYRGSV 179

OY 155 PIPPLTVY 161
| | | | |
Db 180 PIPPLTVY 186

RESULT 18
US-09-900-575-41
; Sequence 41, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burtelin, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 41
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-41

Query Match 41.6%; Score 67; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 2.3e-57;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 IKAGSLI AVLILRQTNNYNSDDFQFWMNIYANDVVPPTGCGDV SAR DVTYTL PDYRGSV 154
| | | | |
Db 120 IKAGSLI AVLILRQTNNYNSDDFQFWMNIYANDVVPPTGCGDV SAR DVTYTL PDYRGSV 179

OY 155 PIPPLTVY 161
| | | | |
Db 180 PIPPLTVY 186

RESULT 19
US-09-900-575-42
; Sequence 42, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burtelin, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 42
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-42

Query Match 41.6%; Score 67; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 2.3e-57;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 IKAGSLI AVLILRQTNNYNSDDFQFWMNIYANDVVPPTGCGDV SAR DVTYTL PDYRGSV 154
| | | | |
Db 120 IKAGSLI AVLILRQTNNYNSDDFQFWMNIYANDVVPPTGCGDV SAR DVTYTL PDYRGSV 179

OY 155 PIPPLTVY 161

Db 180 PIPPLTV 186
|||||

RESULT 20
US-09-900-575-45

; Sequence 45, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900, 575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216, 750
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 45
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-45

Query Match
Best Local Similarity 41.6%; Score 67; DB 10; Length 279;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 IKAGSLAVILRQTNNNNSDDFOFWNNIYANNDDVVPYGGCDVSARDVTVTLPDYRGSV 154
|||||

Db 120 IKAGSLAVILRQTNNNNSDDFOFWNNIYANNDDVVPYGGCDVSARDVTVTLPDYRGSV 179
|||||

Qy 155 PIPPLTV 161
|||||

Db 180 PIPPLTV 186

RESULT 21

US-09-900-575-38
; Sequence 38, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900, 575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216, 750
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-38

Query Match
Best Local Similarity 37.3%; Score 60; DB 10; Length 279;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 AVILRLQTNNNNSDDFOFWNNIYANNDDVVPYGGCDVSARDVTVTLPDYRGSV 161
|||||

Db 127 AVILRLQTNNNNSDDFOFWNNIYANNDDVVPYGGCDVSARDVTVTLPDYRGSV 186

RESULT 22

US-09-900-575-33
; Sequence 33, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900, 575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216, 750
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-33

Query Match
Best Local Similarity 34.8%; Score 56; DB 10; Length 279;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 IKAGSLAVILRQTNNNNSDDFOFWNNIYANNDDVVPYGGCDVSARDVTVTLPDY 150
|||||

Db 120 IKAGSLAVILRQTNNNNSDDFOFWNNIYANNDDVVPYGGCDVSARDVTVTLPDY 175
|||||

RESULT 23
US-09-900-575-23

; Sequence 23, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900, 575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216, 750
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-23

Query Match
Best Local Similarity 29.8%; Score 48; DB 10; Length 279;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PYYVNGQNLVNDLSQIFCHNDYPETITDYVTLQSGAYGGLVNFSG 48
|||||

Db 26 PYYVNGQNLVNDLSQIFCHNDYPETITDYVTLQSGAYGGLVNFSG 73
|||||

RESULT 24

US-09-900-575-24
; Sequence 24, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use

```
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
;
Query Match          26.7%; Score 43; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 3.3e-34;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 IKAGSLIAYLILRQTNNNNSDDFQFVWNIYANDVVPYPTGCD 137
DB 120 IKAGSLIAYLILRQTNNNNSDDFQFVWNIYANDVVPYPTGCD 162

RESULT 25
US-09-900-575-31
; Sequence 31, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burteln, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
;
US-09-900-575-31
Query Match          26.7%; Score 43; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 3.3e-34;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 IKAGSLIAYLILRQTNNNNSDDFQFVWNIYANDVVPYPTGCD 137
DB 120 IKAGSLIAYLILRQTNNNNSDDFQFVWNIYANDVVPYPTGCD 162

RESULT 26
US-10-014-882-2
; Sequence 2, Application US/10014882
; Patent No. US20020107384A1
; GENERAL INFORMATION:
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James
; APPLICANT: Donoho, Gregory
; TITLE OF INVENTION: NO. US20020107384A1 Human Kinase and Polynucleotides Encoding t
; FILE REFERENCE: LEX-0279-USA
; CURRENT APPLICATION NUMBER: US/10/014,882
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: US 60/254,744
; PRIOR FILING DATE: 2000-12-11
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1036
; TYPE: PRT
```

```
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1036)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-014-882-2

Query Match          5.0%; Score 8; DB 12; Length 1036;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 TPVSSAGG 92
DB 12 TPVSSAGG 19

RESULT 27
US-09-901-436A-11
; Sequence 11, Application US/09901436A
; Patent No. US20020098515A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Cyclostatin I
; FILE REFERENCE: PF175D4
; CURRENT APPLICATION NUMBER: US/09/901,436A
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 09/361,737
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: 09/023,073
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 08/470,298
; PRIOR FILING DATE: 1995-06-06
; PRIOR APPLICATION NUMBER: 08/409,731
; PRIOR FILING DATE: 1995-03-24
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 132
; TYPE: PRT
; ORGANISM: human
;
US-09-901-436A-11
Query Match          4.3%; Score 7; DB 10; Length 132;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 VTLOKGS 37
DB 85 VTLOKGS 91

RESULT 28
US-09-912-020-282
; Sequence 282, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlson, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Fioelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DVI
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
```

NUMBER OF SEQ ID NOS: 485
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 282
LENGTH: 1040
TYPE: PRT
ORGANISM: E. coli
US-09-912-020-282

Query Match
Best Local Similarity 4.3%; Score 7; DB 10; Length 1040;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 SLIAYLI 105
|||||
DB 450 SLIAYLI 456

RESULT 29
US-09-815-242-10203

Sequence 10203, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlson, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 10203

LENGTH: 1040

TYPE: PRT

ORGANISM: Escherichia coli

US-09-815-242-10203

QY 99 SLIAYLI 105
|||||
DB 450 SLIAYLI 456

RESULT 30
US-09-864-761-37358

Sequence 37358, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

SEQ ID NO 37358

LENGTH: 40

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AC009721.9

OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.5

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.1

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.3

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.9

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.7

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.2

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.5

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8

QY 33 LORGSA 38
|||||
DB 7 LORGSA 12

RESULT 31
US-09-864-761-37358

Sequence 37358, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

```
US-09-726-643-65
; Sequence 65, Application US/09726643
; Patent No. US20020028449A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/09/726,643
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-726-643-65

Query Match
Best Local Similarity 3.7%; Score 6; DB 10; Length 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 VTWTLP 148
Db 3 VTWTLP 8
|||||

RESULT 32
US-09-864-761-34310
; Sequence 34310, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34310
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009238.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.73
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: SWISSPROT HIT: Q59050, EVALU6 6.00e-07
; OTHER INFORMATION: EST_HUMAN HIT: A1884989.1, EVALU6 3.00e-26
US-09-864-761-34310

Query Match
Best Local Similarity 3.7%; Score 6; DB 10; Length 54;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 DVSARD 142
Db 16 DVSARD 21
|||||

RESULT 33
US-09-864-761-35614
; Sequence 35614, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
```

;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 35614
;; LENGTH: 62
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO 298044.10
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
;; OTHER INFORMATION: SWISSPROT HIT: Q95479, EVALUATE 7.00e-30
;; OTHER INFORMATION: EST_HUMAN HIT: AWS03925.1, EVALUATE 6.00e-29
US-09-864-761-35614

Query Match
Best Local Similarity 3.7%; Score 6; DB 10; Length 62;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 LONGSA 38
|11111|
Db 27 LONGSA 32

RESULT 34
US-09-867-550-1142
;; Sequence 1142, Application US/09867550
;; Patent No. US20020082206A1
;; GENERAL INFORMATION:
;; APPLICANT: Leach, Martin D.
;; APPLICANT: Mehreban, Fuad,
;; APPLICANT: Conley, Pamela
;; APPLICANT: Law, Debbie
;; APPLICANT: Topper, James
;; TITLE OF INVENTION: Thereby
;; FILE REFERENCE: 21402-013 (Cura-313)
;; CURRENT APPLICATION NUMBER: US/09/867,550
;; PRIOR FILING DATE: 2001-09-20
;; PRIOR APPLICATION NUMBER: USSN 60/208,427
;; PRIOR FILING DATE: 2000-05-30
;; NUMBER OF SEQ ID NOS: 2125
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 1142
;; LENGTH: 93
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-867-550-1142

Query Match
Best Local Similarity 3.7%; Score 6; DB 10; Length 93;
Matches 100.0%; Pred. No. 60;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 VYVTLP 148
|11111|
Db 73 VYVTLP 78

RESULT 35
US-09-864-761-35148
;; Sequence 35148, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharon G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Aecmlca-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; PRIOR FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 35148
;; LENGTH: 99
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL022337.19
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.96
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
US-09-864-761-35148

Query Match 3.7%; Score 6; DB 10; Length 99;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 VGONLV 10
|||||
Db 2 VGONLV 7

RESULT 36
US-09-925-300-1428
Sequence 1428, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruden,
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO. 1428
LENGTH: 112
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (20)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (57)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1428

Query Match 3.7%; Score 6; DB 10; Length 112;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 62 TSETPR 67
|||||
Db 5 TSETPR 10

RESULT 37
US-09-728-914-10
Sequence 10, Application US/09728914
Patent No. US20010046499A1
GENERAL INFORMATION:
APPLICANT: KANTOR, FRED S.
APPLICANT: FIKRIG, EROL
APPLICANT: DAS, SUBRAVA
TITLE OF INVENTION: TICK ANTIGENS AND COMPOSITIONS AND METHODS COMPRISING
TITLE OF INVENTION: THEM
FILE REFERENCE: YU-107
CURRENT APPLICATION NUMBER: US/09/728,914
CURRENT FILING DATE: 2000-12-01
PRIOR APPLICATION NUMBER: 60/169,048
PRIOR FILING DATE: 1999-12-03
PRIOR APPLICATION NUMBER: 60/240,716
PRIOR FILING DATE: 2000-10-16
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 144

TYPE: PRT
ORGANISM: Ixodes scapularis
US-09-728-914-10

Query Match 3.7%; Score 6; DB 10; Length 144;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 144 TVTLPD 149
|||||
Db 106 TVTLPD 111

RESULT 38
US-09-864-761-43168
Sequence 43168, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 43168
LENGTH: 182
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004886.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
OTHER INFORMATION: EST_HUMAN HIT: BE967305.2, EVALU 5.90e+00

OTHER INFORMATION: SWISSPROT HIT: P27110, EVALDE 3.50e-01
US-09-864-761-43168

Query Match
Best Local Similarity 3.7%; Score 6; DB 10; Length 182;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 VVDLST 15
|||||
DB 47 VVDLST 52

RESULT 39
US-09-739-907-65

Sequence 65, Application US/09739907
Patent No. US20010012889A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 36 Human Secreted Proteins
FILE REFERENCE: P2022P1
CURRENT APPLICATION NUMBER: US/09/739,907
CURRENT FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: 09/348,457
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: 60/070,567
PRIOR FILING DATE: 1998-01-07
PRIOR APPLICATION NUMBER: 60/070,692
PRIOR FILING DATE: 1998-01-07
PRIOR APPLICATION NUMBER: 60/070,704
PRIOR FILING DATE: 1998-01-07
PRIOR APPLICATION NUMBER: 60/070,658
PRIOR FILING DATE: 1998-01-07
NUMBER OF SEQ ID NOS: 196
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 65
LENGTH: 183
TYPE: PRT
ORGANISM: Homo sapiens
US-09-739-907-65

Query Match
Best Local Similarity 3.7%; Score 6; DB 10; Length 183;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 LYLPV 87
|||||
DB 17 LYLPV 22

RESULT 40
US-10-036-869-36

Sequence 36, Application US/10036869
Patent No. US20020151516A1
GENERAL INFORMATION:

APPLICANT: Mixson, James A
TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
THERAPY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:

ADDRESSEE: Connolly, Boye, Lodge, & Hultz
STREET: 1220 Market Street, P.O. Box 2207
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19899

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/036,869

FILING DATE: 29-NO. US20020151516A1-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/985,526

FILING DATE: <Unknown>

APPLICATION NUMBER: US 08/608,845

FILING DATE: 16-JUL-1996

ATTORNEY/AGENT INFORMATION:

NAME: McMorrow Jr., Robert G

TELECOMMUNICATION INFORMATION:

TELEPHONE: (302) 658-9141

TELEFAX: (302) 658-5613

SEQUENCE CHARACTERISTICS:

LENGTH: 185 amino acids

TYPE: amino acid

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 36;

US-10-036-869-36
QY 151 RGSVPI 156
|||||
DB 66 RGSVPI 71

RESULT 41
US-09-998-831-13

Sequence 13, Application US/09998831
Patent No. US20020119153A1
GENERAL INFORMATION:

APPLICANT: Philip E. Thorpe
TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
INHIBITING VEGF
FILE REFERENCE: 4001.002584
CURRENT APPLICATION NUMBER: US/09/998,831
CURRENT FILING DATE: 2001-11-30
PRIOR APPLICATION NUMBER: 09/561,108
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 191
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-998-831-13

Query Match
Best Local Similarity 3.7%; Score 6; DB 10; Length 191;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 RGSVPI 156
|||||
DB 73 RGSVPI 78

RESULT 42
US-09-925-302-728

Sequence 728, Application US/09925302
Patent No. US20020044941A1
GENERAL INFORMATION:

APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P4104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918

```

; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 728
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-302-728

Query Match          3.7%; Score 6; DB 10; Length 192;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 VLSNFS 47
|111111
DB 132 VLSNFS 137

RESULT 43
US-09-734-017A-74
; Sequence 74, Application US/09734017A
; Patent No. US2002014242A1
; GENERAL INFORMATION:
; APPLICANT: Lerchl, Jens
; APPLICANT: Renz, Andreas
; APPLICANT: Ehrhardt, Thomas
; APPLICANT: Reindl, Andreas
; APPLICANT: Clippus, Petra
; APPLICANT: Bischoff, Friedrich
; APPLICANT: Frank, Markus
; APPLICANT: Freund, Annette
; APPLICANT: Duwenig, Elke
; APPLICANT: Schmidt, Ralf-Michael
; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved
; TITLE OF INVENTION: the
; TITLE OF INVENTION: synthesis of amino acids, vitamins, cofactors, nucleosides and
; TITLE OF INVENTION: nucleosides
; FILE REFERENCE: BASF-NAE-1331-99-US
; CURRENT APPLICATION NUMBER: US/09/734,017A
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/171,100
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.1/Mordperfect
; SEQ ID NO 74
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Physcomitrella patens
US-09-734-017A-74

Query Match          3.7%; Score 6; DB 10; Length 194;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 SAGGLV 94
|111111
DB 36 SAGGLV 41

RESULT 44
US-09-815-242-13993
; Sequence 13993, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

```

```

; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13993
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(213)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13993

Query Match          3.7%; Score 6; DB 10; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 VTLPDY 150
|111111
DB 26 VTLPDY 31

RESULT 45
US-09-946-239-10
; Sequence 10, Application US/09946239
; Patent No. US20020044945A1
; GENERAL INFORMATION:
; APPLICANT: Barr, Margaret C.
; TITLE OF INVENTION: Polypeptide Sequences
; FILE REFERENCE: 18617.0059
; CURRENT APPLICATION NUMBER: US/09/946,239
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: US 09/238,303, US 60/072,927
; PRIOR FILING DATE: 1999-01-28, 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 10
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: protein encoded by the vlf gene of a recombinant viral
; OTHER INFORMATION: clone constructed from the genomic DNA of a Pallas's cat felin
; OTHER INFORMATION: Immunodeficiency virus
US-09-946-239-10

Query Match          3.7%; Score 6; DB 10; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 GLVKA 97
|111111
DB 154 GLVKA 159

```

Tue Dec 3 12:14:25 2002

us-09-900-575-29_copy_26_186.oligo.rpb

Page 21

Search completed: November 28, 2002, 19:09:50
Job time : 16 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 28, 2002, 19:09:06 ; Search time 1977 Seconds

(without alignments)
1318.904 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 848
Sequence: 1 PVPVWGONLVLDISTQIFCH.....DVTVLPDYGVSVPPLTVY 161

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xih
-O=/gen2_1/USPRO.spool/US09900575/runat.22112002_130709_4559/app_query.fasta.1.327
-DB=EST -OFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=Diosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09900575_QCGN.1.1.899_@runat.22112002_130709_4559 -NCPU=6 -ICPU=3
-NO_XLPY -NO_MMAP -LARGEOUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_estc3:*
13: gb_estc4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	269.5	31.8	914	14 B0143831	B0143831 NF003F10DT
C 2	174.5	20.6	256	10 BE344026	BE344026 EST405104
C 3	89.5	10.6	864	17 CNS06R7	AL411881 T3 end of
C 4	85.5	10.1	659	17 AZ368147	AZ368147 243PvF08
C 5	85.5	10.1	802	12 BG666096	BG666096 DRACN06
C 6	84	9.9	510	13 B1624785	B1624785 RH64464.5
C 7	84	9.9	720	14 B0968753	B0968753 QH835A23
C 8	83.5	9.8	653	13 B1293068	B1293068 B1293068
C 9	83.5	9.8	807	12 BF649263	BF649263 NF055G05E
C 10	83.5	9.8	807	12 BG708796	BG708796 602673080
C 11	83.5	9.8	834	12 BG621049	BG621049 602616925
C 12	83.5	9.8	911	17 BH147842	BH147842 ENTPRS5TR
C 13	83	9.8	756	14 BQ970561	BQ970561 QH842H01
C 14	82.5	9.7	413	12 BF557712	BF557712 UI-R-CO-h
C 15	82.5	9.7	501	12 BG664251	BG664251 DRABRC11
C 16	82.5	9.7	518	12 BG664658	BG664658 DRABRC11
C 17	82.5	9.7	530	12 BF072843	BF072843 DRABRC11
C 18	82.5	9.7	537	12 BG670398	BG670398 DRABRC11
C 19	82.5	9.7	573	12 BG671004	BG671004 DRABRC10
C 20	82.5	9.7	602	13 B1797757	B1797757 H086C03 E
C 21	82.5	9.7	696	12 BG665535	BG665535 DRACEC11
C 22	82.5	9.7	697	12 BG669393	BG669393 DRACNAF07
C 23	82.5	9.7	721	10 BB355470	BB355470 BB355470
C 24	82.5	9.7	725	10 AW914983	AW914983 EST346287
C 25	82.5	9.7	747	12 BG673676	BG673676 DRNB2B02
C 26	82.5	9.7	789	12 BG671529	BG671529 DRNB2G05
C 27	82.5	9.7	1312	13 B1079789	B1079789 602876103
C 28	82	9.7	581	10 AV986132	AV986132 AV986132
C 29	82	9.7	582	10 AV996881	AV996881 AV996881
C 30	82	9.7	557	10 AV984947	AV984947 AV984947
C 31	82	9.7	614	10 AV996923	AV996923 AV996923
C 32	82	9.7	642	10 AV981815	AV981815 AV981815
C 33	82	9.7	647	10 AV974294	AV974294 AV974294
C 34	82	9.7	651	10 AV673892	AV673892 AV673892
C 35	82	9.7	689	10 AV970876	AV970876 AV970876
C 36	82	9.7	939	9 AL667637	AL667637 AL667637
C 37	81.5	9.6	544	12 BG666005	BG666005 DRACMA09
C 38	81	9.6	443	9 AA969177	AA969177 OP58C03.s
C 39	81	9.6	594	17 BH519770	BH519770 BOCDB45TF
C 40	81	9.6	669	17 AG072173	AG072173 Pan trogl
C 41	81	9.6	826	9 AL666950	AL666950 AL666950
C 42	81	9.6	868	9 AL667437	AL667437 AL667437
C 43	80.5	9.5	675	13 B1208686	B1208686 EST526726
C 44	80.5	9.5	731	13 B1208424	B1208424 EST526644
C 45	80.5	9.5	779	12 BG428847	BG428847 602494632

ALIGNMENTS

RESULT 1
LOCUS B0143831/c 914 bp mRNA linear EST 24-APR-2002
DEFINITION NF003F10DT.F1079 Drought Medicago truncatula cDNA clone NF003F10DT
5', RNA sequence.
ACCESSION B0143831
VERSION B0143831.1 GI:20280890
KEYWORDS EST.
SOURCE Medicago truncatula
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
REFERENCE
1 (bases 1 to 914)

AUTHORS		Torrez-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Imman, J.T., Weller, J.W. and May G.D.			
TITLE		Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula drought library			
JOURNAL		Unpublished (2000)			
COMMENT		Contact: May GD Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA Tel: 580 221 7391 Fax: 580 221 7380 Email: gdmay@noble.org Insert Length: 914 Std Error: 0.00 Plate: 003 row: F column: 10 Seq primer: TCACACGGAACAGCATATGAC.			
FEATURES		Location/Qualifiers			
SOURCE		1..914			
		/organism="Medicago truncatula"			
		/db_xref="taxon:3880"			
		/clone="NF003F10DT"			
		/clone_lib="Drought"			
		/tissue_type="plantlets"			
		/dev_stage="pooled timepoints"			
		/note="Vector: Lambda Zap; Contains a mixture of entire platelets harvested in a series of days-post-watering timepoints."			
BASE COUNT		257 a 200 c 193 g 264 t			
ORIGIN					
Alignment Scores:					
Pred. No.:		1.09e-21		Length: 914	
Score:		269.50		Matches: 53	
Percent Similarity:		65.00%		Conservative: 25	
Best Local Similarity:		44.17%		Mismatches: 41	
Query Match:		31.78%		Indels: 1	
DB:		14		Gaps: 1	
US-09-900-575-29_COPY_26_186 (1-161) x BQ143831 (1-914)					
QY	39	TYRGYGLYVALLSERASNPHERSGLYTHRALVSYRSERGLYSERTYRPRO	58		
DB	459	TACGGGCGCCCTCCATACAGCATATTAAGATCGCTGATCGAATATGATGACCTACCC	400		
QY	59	PHERTHTHSESCILNTHPRNRYVALVLYRAENSERATGTHTASPLSPROTTP	78		
DB	399	TTCCCGTTGACGACAATATCAACGATTGGATTGGTATAGACCCGATGCCATTA	340		
QY	79	PROVALALALEYRLEUTHPRPROVALSERASRLAGLYLEUVALLELYSALAGLY	98		
DB	339	CCGTTAAACTGTATATATACCCAGTGGGGCGCGGAGTGTATATAAAGCGGG	280		
QY	99	SERLEUILLALAVALLEULEARGLNTHPRASNASTRYASNSER---ASPLSPHE	117		
DB	279	GAAGTCATTGCCCGTATTCATATGATATAAATTCGAACATTAGGAAGCGGAATCCGGCT	220		
QY	118	GLNPHEVALTPASNLLETYRALASNASASPAVALVALPPTHGSLGYLCYSASP	137		
DB	219	AATTTTACTGGATATTTATTCAAATATACATGTTGTATGCCAAGCGGCTGCACC	160		
QY	138	VALSERALARQSPVALTHVALTHRLSEUPROASPTYRAGLYSERVALPROLLEPRO	157		
DB	159	GTTGATGATCGAATATCTACCGGTGATCTCCCGACTTCTCTGGAAGCCAGAAATTCCT	100		
RESULT 2					
LOCUS		BE434026		256 bp mRNA linear EST 18-MAY-2001	
DEFINITION		EST4051010 tomato breaker fruit1, t1GR Lycopersicon esculentum cDNA			
ACCESSION		BE434026			
VERSION		BE434026.1 GI:9431869			
KEYWORDS		EST.			
SOURCE		Tomato.			
ORGANISM		Lycopersicon esculentum			

REFERENCE	TITLE	FEATURES
ATTORNS	JOURNAL	source
REFERENCE	COMMENT	Location/Qualifiers
1 (bases 1 to 256)	Generation of ESTs from tomato fruit tissue, breaker stage	1..256
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:	Unpublished (2000)	/organism="Lycopersicon esculentum"
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;	Contact: CUGI	/cultivar="TA496"
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;	Clemson University Genomics Institute	/db_xref="taxon:4081"
Lycopersicon.	Clemson University	/clone="CLEG13H23"
1 (bases 1 to 256)	100 Jordan Hall, Clemson, SC 29634, USA	/clone.lib="tomato breaker fruit, TIGR"
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:	Email: http://www.genome.clemson.edu/orders/index.html	/tissue-type="pericarp"
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;	5 prime sequence.	/dev_stage="breaker"
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		/lab_host="SOLR"
Lycopersicon.		/note="Vector: plasmidscriptsmcUnadapt; Site_1: EcoRI;
1 (bases 1 to 256)		Site_2: XhoI; Fruit were harvested at the breaker stage
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		(first sign of lycopene accumulation on the blossom end
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		the fruit). Fruit were cut in half and the seeds and
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		locules were discarded prior to freezing the pericarp."
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots;		
Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
Lycopersicon.		
1 (bases 1 to 256)		
Alcayota, Viridiplantae: Streptophyta: Embryophyta: Tr		

Whatman CFI1 powder (1:2 ratio volume of blood to CFI1), and finally centrifuged through a 50% Percoll density cushion. Purified DNA was digested with mung bean nuclease in the presence of 44% formamide at 50°C as described (Vernick, K.D., Imberski, R.B., and McCutchan, T.F., 1988. Nucleic Acids Research 16:6883-6896). Digested DNA was blunt-ended using T4 DNA polymerase and size fractionated over a Sepharose CL-2B column. Fractions in the size range 500bp-4kb were ligated into the Eco RV site of pBluescript SK(+), and E. coli XL-10 Gold transformed with the ligation mixture."

BASE COUNT 122 a 163 c 148 g 225 t 1 others
ORIGIN

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score: 7.47	659	32	18	32	34	6
Percent Similarity: 43.10%						
Best Local Similarity: 27.59%						
Query Match: 10.08%						

US-09-900-575-29_COPY_26_186 (1-161) x AZ568147 (1-659)

```

OY 9 LeuValValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr 28
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 130 TTGTTAAAGCAGCGAGTTGTTCAAGTGTCTCCGGGAATTCATCGCTTGTACT 189
OY 29 -----AspTyrValThrLeuGlnArgly----- 36
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 190 AAGCGGATTTGGTTCTTCATCCAGGGGATGTACCGATAGGCGCTTTCTATCCGT 249
OY 37 -----SerAlaTyrIleGlyValLeuSerAsnPhSerGlyThrValIleTyrSer 53
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 250 TTCTGCGCGTCCGACTGTGCTGGGCT--TCCAATTTTGTTGGGCTTCTCTCTTCT 306
OY 54 GlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsnSerArg 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 307 GGGGACCCCTACCTTTCATGCTTGTGCGCTTCTGCCCCCAATGT----- 351
OY 74 ThrAspLysProThrProValAlaLeuTyrLeuThr-ProVal-----Se 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 352 -----CCTTTTCCCATTTGGACTTGTGCACATCTTATGAACAGTGTGAAGAA 402
OY 88 rSerAlaGlyGlyLeu-----ValIleLysAlaGlySer 99
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 403 CTCTCTAGGGGCGATCACCAGCAGAGAAATTAAATGGTTGG 448

```

RESULT 5
LOCUS BG666096 802 bp mRNA linear EST 30-APR-2001
DEFINITION DRACNG06 Rat DRG Library Rattus norvegicus cDNA clone DRACNG06 5',
mRNA sequence.
ACCESSION BG666096
VERSION BG666096.1 GI:13888018
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.,
Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang
,X., Chen,Z., Han,Z.G. and Zhang,X.
Identification of gene expression profile of dorsal root ganglion
in the rat peripheral axotomy model of neuropathic pain
Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)

TITLE
JOURNAL MEDLINE
COMMENT
Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China

Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome
Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
Pudong New Area, P.R.China. please contact with Zhang Xu
(xu.zhang@ion.ac.cn) or Han Zeyang (hanzg@chgc.sh.cn)
PCR primers
FORWARD: T7
BACKWARD: T7
Seq primer: T7
POLYA=No.

FEATURES
SOURCE Location/Qualifiers

1..802
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRACNG06"
/clone_1lb="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/note="Total RNA was isolated from hypothalamus and
transcribed into cDNA, which was then used as template in
PCR. The PCR products were separated on sequencing gel.
The differential bands were cut, reamplified, cloned into
pMD18-T vector and confirmed by Northern blot."

BASE COUNT 204 a 216 c 193 g 189 t

ORIGIN

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score: 10.1	802	29	18	42	37	3
Percent Similarity: 37.30%						
Best Local Similarity: 23.02%						
Query Match: 10.08%						

US-09-900-575-29_COPY_26_186 (1-161) x BG666096 (1-802)

```

OY 13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 420 TTACAACTGCTTGTCTGTGTGCACCAACGTGCGCGGAAGTTAGACC----- 467
OY 33 LeuGlnArgGlySerAlaTyrIleGlyValLeuSerAsnPhSerGlyThrValIleTyr 52
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 467 ----- 467
OY 53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg-----Val 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 468 -----TATCACCACACACCTCCACACAGCCCGCAAAAGACAGCCCACTCAA 518
OY 69 ValTyrAsnSerArgThrAspLysProThrProValAlaLeuTyrLeuThrProValSer 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 519 GTCTCTTCTTCCACAGTGAACCGGTGTGTACAGTACGATTTTGGACACACAGCT 578
OY 89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 579 CTAACGTGTGAACGATTTGATCTGCACACCTAAGTATATGATCATCTCACTCAGCAGCA 638
OY 109 ThrAsnAspTyrAsnSerAspPheGlnPheValThrPsnIleTyrAlaAsnAsp 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 639 GAATCAAAATTTGGCTCACTAGACGCTCTCTTT-----CAGGATGAC 680
OY 129 ValValValProThrGly 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 681 ACCATTGTCCCATAGA 698

```

RESULT 6
LOCUS B1624785 510 bp mRNA linear EST 07-SEP-2001
DEFINITION RH64464.5prine RH Drosophila melanogaster normalized Head p1c-1
Drosophila melanogaster cDNA clone RH64464 5 similar to Glyp:

FBan007254 GO: [glycogen phosphorylase (GO:0008184): enzyme (GO:0003824): phosphorylase (GO:0004645)] located on: 2L 22C3-22C3
 : : 08/23/2001, mRNA sequence.
 ACCESSION B1624785
 VERSION B1624785.1 GI:15520310
 KEYWORDS EST.
 SOURCE Fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 510)
 Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Mista, S., Munhall, C. J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phonsavanong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and Rubin, G. M.
 BDGP/HMT RH Drosophila EST Project
 Unpublished (2001)
 COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: <http://www.fruitfly.org/EST/estefruitfly.berkeley.edu>
 hit genomic AB03584: arm: 2L 1824960, 2149443
 estimated-cyto: 22B4-22D2: 08/23/2001
 Plate: RH 644 row: F column: 4
 High quality sequence stop: 332.
 Location/Qualifiers
 1..510
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="RH64464"
 /clone_lib="RH Drosophila melanogaster normalized Head
 pRc-1"
 /sex="male and female"
 /dev_stage="Adult"
 /lab_host="DHS-alpha Tona"
 /note="Organ: head; Vector: pRc1; Site: 1: XhoI; Site: 2: BamHI. Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
 BASE COUNT 145 a 142 c 135 g 88 t
 ORIGIN

Alignment Scores:
 Pred. No.: 7.59 Length: 510
 Score: 84.00 Matches: 29
 Percent Similarity: 47.62% Conservative: 11
 Best Local Similarity: 34.52% Mismatches: 37
 Query Match: 9.91% Indels: 7
 DB: 13 Gaps: 2

US-09-900-575-29_COPY_26_186 (1-161) x B1624785 (1-510)

OY 53 Ser1ySer1yrrProphrProthThrSerclunThrProArg_ValValTyraanse 72
 Db 505 GCAGGTTCTGCATATCCATCACTGATGATGCGCTCGCATCTCGATATTC 446

OY 72 rArGThAsp1yrrProthrProvalAlaLeuThrPleuthrProvalSerSerla1ygl 92
 Db 445 ACAGCTGATCAT-----GGTGTGCTCAGTACAGCGACCCAGTACTCCAGCGACA 392

OY 92 yLeuVal1le1yAlaGlySerLeu1leAlaVal1leuAlaGlnThraSnaSnty 112
 Db 391 GATAGTAGCCGCGCTTGATCTCTCGTAGAGTACTGCTCTCGTACG-----343

OY 112 rAnSerAspAspPheGlnPheVal1rPasn1leTyra1aAsnAspAlaValValP 132
 Db 342 -ATTCATCGGCCACCACTGTTGCTCTGACGGGTGGCGACGGTGAAGTACTGCC 284

OY 132 oThrGlygly 135
 Db 283 TCAGGTCGA 274

RESULT 7
 LOCUS B0968753
 DEFINITION QHB35A23.yg.ab1 QH-ABCD1 sunflower RHAB01 Helianthus annuus cDNA
 clone QHB35A23, mRNA sequence.
 ACCESSION B0968753
 VERSION B0968753.1 GI:22386274
 KEYWORDS EST.
 SOURCE common sunflower.
 ORGANISM Helianthus annuus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae; Heliantheae; Helianthus.
 1 (bases 1 to 720)
 Kozik, A., Michelmore, R. W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolman, J., Slabaugh, M. S., Livingston, K., Zhou, Y., Lai, E., Church, S., Jackson, L. and Bradford, K.
 lettuce and Sunflower ESTs from the Composite Genome Project
 http://compomics.ucdavis.edu/
 Unpublished (2002)
 COMMENT Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Asmundson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig QH-CA-Contig1697, see <http://cgpb.ucdavis.edu/>
 for details.
 Plate: QHB35 row: A column: 23.
 Location/Qualifiers
 1..720
 /organism="Helianthus annuus"
 /cultivar="RHAB01"
 /db_xref="taxon:4232"
 /clone="QHB35A23"
 /clone_lib="QH-ABCD1 sunflower RHAB01"
 /lab_host="E.coli"
 /note="Vector: pBRCDNA51AB; The library was constructed from 11 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpb.ucdavis.edu/>
 TAG_LIB=QH-ABCD1 sunflower RHAB01
 TAG_TISSUE=chemical induction
 TAG_SEQ=GTGACCGGCGG"
 BASE COUNT 206 a 140 c 164 g 210 t
 ORIGIN

Alignment Scores:
 Pred. No.: 13 Length: 720
 Score: 84.00 Matches: 35
 Percent Similarity: 43.48% Conservative: 25
 Best Local Similarity: 25.36% Mismatches: 56
 Query Match: 9.91% Indels: 22
 DB: 14 Gaps: 6

US-09-900-575-29_COPY_26_186 (1-161) x B0968753 (1-720)

OY 30 TyrValThrLeuGlnArgGlySer1aTyrglyGlyVal1leuSerAsnPhSerGlyThr 49
 Db 16 TGCTACTGATTAACGAGCTTGCTTATATATGATGATGATCTCTCGGCTTAT 75

ORIGIN	BASE COUNT
OY 50	162 a 180 c 203 g 108 t
50	
VallystYrSerGlySerSerTytrProPheProthrHrHrSerGlyuHrProArGValVal	
69	
Db 76	
GGAAATGACACG---CCTCAATATTAATG---GTCAAGACGACGCTGCCCAAGTG---	
126	
OY 70	
TyrAnsSerArGthrAspLysProTirProValAlaLeuTyrlleuThrProValSerSer	
89	
Db 127	
-----AMCGAACCGACACACCATGTTACTGTTAGTGATGATGCATTCGATTTCTATAGT	
180	
OY 90	
AlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValIleLeuArgGlnThr	
109	
Db 181	
AGCATATGATCATCATTTATATGGAAGGTAAACCATGACATGATGATGATACGAACAATATTTT	
240	
OY 110	
AsnAsenTyraSnSerAsp---AspPheGlnPheValTyrlPasnIleTyrlAlaAsnAsnAsp	
128	
Db 241	
GTTCAGTCAACAAAGTTGATGTTGTCTCTCGACACGTTACGCTTACGACGCAACGAC	
300	
OY 129	
ValVal-----ValProthrGlyGlyGlyAspValSerAlaArgasp	
142	
Db 301	
CGGGATGCAATATAGTCTTACCAATATGCAAAAGGGTTGTGCAT-----	
345	
OY 143	
ValThrValThrLeuProAspTyrArgLysSerValProIleProLeuThrVal	
160	
Db 346	
-----CCTAAATATGATGATGATCTGCCCTGTTTACATAAACGTT	
384	
RESULT 8	
Bj293068/c	653 bp mRNA linear EST 09-APR-2002
LOCUS	Bj293068 Y. Ogihara unpublished cDNA library, Wh.SL Triticum
DEFINITION	aestivum cDNA clone whsl30h21 5', mRNA sequence.
ACCESSION	Bj293068
VERSION	Bj293068.1 GI:20108952
KEYWORDS	EST.
SOURCE	bread wheat.
ORGANISM	Triticum aestivum
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
TITLE	; Triticeae; Triticum.
JOURNAL	1 (bases 1 to 653)
COMMENT	Ogihara,Y. and Murai,K. Expressed genes in Triticum aestivum Unpublished (2002) Contact: Tadasu Shin-I Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855. Email: tshinigenes.nig.ac.jp. Location/Qualifiers 1. .653 /organism="Triticum aestivum" /cultivar="Chinese Spring" /db_xref="taxon:4565" /clone="whsl30h21" /clone_1fb="Y. Ogihara unpublished cDNA library, Wh.SL" /tissue_type="seed DPA30" /dev_stage="Flekes" scale 11.3" /note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown under hydroponic conditions at UC Davis, salt stressed for 12 hours, and for 7 days, then dissected and frozen (Akhunov in J Dvorak lab). Total RNA was prepared from sheath tissue, equal quantities of RNA were pooled from the two samples, polyA was purified from the pooled RNA, a cDNA library was made, and the cDNA clones were in vivo excised to give plusescript phagemids in the JI Close lab at the University of California, Riverside (Akhunov, Chih, Choi, Close, Fenton, Kiantian, Otto, Simons, Zhang). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

Alignment Scores:	12.8	Length:	653
Pred. NO.:	83.50	Matches:	24
Score:	46.27%	Conservative:	7
Percent Similarity:	35.82%	Mismatches:	17
Best Local Similarity:	9.85%	Indels:	19
Query Match:	13	Gaps:	3
DB:			
US-09-900-575-29_COPY_26_186 (1-161) x BU293068 (1-653)			
QY	43	LeuSerAsnPhseSerGlyThrVallystrSerGlySerSerTyProPheProThr 62	
Db	252	CTCAGTGGCTCCGTAGAGCCGCCGCGGTGGTCCCGCTGACCTGACCCGACAGT 193	
QY	63	SerGluThrProAlaValTyTrasAnSerArGthrAspIysProTTPProValAlaLeu 82	
Db	192	GCCAGTCGCCCCGTGA-----CCCTGGCCCGTAGCC-- 160	
QY	83	TyrLeuThrProValSerSerAlaGlyIy-----LeuAlIleAlaGlySer 99	
Db	159	-----GCCGGGCGCGTGTGCTGCTCTTGTTACGCCGGGGAG 121	
QY	100	LeuIleAlaValLeuIleLeu 106	
Db	120	CTTCTCTTGATCTTGCTT 100	
RESULT 9			
BF649263			
LOCUS	BF649263	657 bp	mRNA linear EST 20-DEC-2000
DEFINITION	BF649263	clone NF05G05EC1F1038	elicited cell culture Medicago truncatula cDNA
ACCESSION	BF649263		
VERSION	BF649263.1	GI:11914393	
KEYWORDS	EST.		
SOURCE	barrel medic.		
ORGANISM	Medicago truncatula		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.		
REFERENCE	1 (bases 1 to 657)		
AUTHORS	Torres-Terrez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Imman, J.T., Wellner, J.W. and May, G.D.		
TITLE	Expressed Sequence tags from the Samuel Roberts Noble Foundation - Center for Medicago Genomics Research		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Dixon RA Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA Tel: 580 221 7302 Fax: 580 221 7380 Email: radixon@noble.org Insert Length: 657 Std Error: 0.00 Plate: 055 row: G column: 05 Seq primer: TCACACGAGAAACAGCTATGAC.		
FEATURES			
Source			
	1..657		
	/organism="Medicago truncatula"		
	/db_xref="taxon:3880"		
	/clone="NF05G05EC"		
	/clone_lib="Elicited cell culture"		
	/tissue_type="Cell cultures derived from root tissues"		
	/dev_stage="Cell suspensions were subcultured every 14 days. Cells were induced six days after subculture"		
	/note="Vector: Lambda Zap; Cells were induced with yeast cell wall extracts equivalent to 50ug/ml glucose in the final concentration. Samples were taken at 0.5, 1, 12 and 24 hours after induction. Equal amounts of RNA from each time point were pooled and used for mRNA isolation."		
BASE COUNT	205 a 116 c 124 g 210 t		2 others
ORIGIN			

Alignment Scores:

Pred. No.:	12.9	Length:	657
Score:	83.50	Matches:	34
Percent Similarity:	43.22%	Conservative:	17
Best Local Similarity:	28.81%	Mismatches:	30
Query Match:	9.85%	Indels:	37
DB:	12	Gaps:	7

US-09-900-575-29_COPY_26_186 (1-161) x BF649263 (1-657)

QY 2 ValValAsnValGlyClnAsnLeuValVal-----AspLeu 13
 :::|||||:::|||||
DB 349 GCGATAAATATGCGTGAATCTGCAATCATTTGATGAGTTGCATATCAATTTTA 408
 |||||
QY 14 SerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeu 33
 |||||:::|||||:::|||||
DB 409 TCAACA-----GTGCAACATGATGATTAATGTTGAACATTATGTTGAT----- 450
 |||
QY 34 GlnATGlySerAlaTyrGlyValLeuSerAsnPheserGlyThrValLysTyrSer 53
 |||:::|||||
DB 451 -----TACAAAGATAACTTTCATCTTATCAAGCATTTGATGTGATTTCT 495
 |||
QY 54 GlySerSerTyrPro-----PheProThrThrSerGluThrProArg----- 67
 |||:::|||||:::|||||
DB 496 GGAAGTAGTGTCCAGATTGTTGATGATCAAGACAAACAGACAAACAGATGAC 555
 |||:::|||||
QY 68 ---ValValItyrAsnSerArgThrAspLysProTyrProValAlaLeuTyrLeuThrPro 86
 |||:::|||||
DB 556 ATGATTTGATCTCTCT-----CCACTTCATCTCTCCCT 591
 |||
QY 87 ValSerSerAlaGlyClnLeuValIle-----LysAlaGlySerLeuIle 101
 |||:::|||||
DB 592 TTGNTGGGTTTGGTTTGGTTTATCTTCCTTCAGAACGAGGAATATGATA 645
 |||:::|||||

RESULT 10

LOCUS BG708796 807 bp mRNA linear EST 07-MAY-2001
DEFINITION 602673080F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4796010 5',
 mRNA sequence.
ACCESSION BG708796
VERSION BG708796.1 GI:13986493
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
 Toshlyuk and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LMA10679 row: O column: 19
 High quality sequence stop: 760.
 Location/Qualifiers
 1. 807

FEATURES
SOURCE /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4796010"
 /clone_id="NIH_MGC_96"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /notes="Organ: brain; Vector: pBluescriptR (modified
 pBluescript KS+); Site:1; BamHI: Site2; SalI:XhoI (gtcgag
); Oligo-dT primed using primer 5'-tttttttttttttttTA-3',

size-selected for average insert size 2.3 kb and
normalized to ROP 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH-MGC Library."
BASE COUNT 171 a 211 c 223 g 202 t
ORIGIN

Alignment Scores:

Pred. No.:	17.8	Length:	807
Score:	83.50	Matches:	46
Percent Similarity:	42.95%	Conservative:	21
Best Local Similarity:	29.49%	Mismatches:	44
Query Match:	9.85%	Indels:	45
DB:	12	Gaps:	8

US-09-900-575-29_COPY_26_186 (1-161) x BG708796 (1-807)

QY 31 ValThrLeuGlnArgGly-----SerAlaTyrGlyGly-----ValLeuSerAsnPh 46
 |||||:::|||||
DB 773 GTACGGTGAATAATGAGCATCTACAGATGAGGTGAGGAGAAATCTGCCGTCAC 714
 |||||
QY 47 SerGlyThrValLysTyrSerGlySerSerTyrProPheProThrThr----- 62
 |||||:::|||||
DB 713 TCAGAACAGCTTAATAGATCAAAAGCCGATCCCAAGAGTACACGATGTAAATCAAT 654
 |||
QY 63 -----SerGluThr-ProArgValValTyrAsnSerArgThrAspLysProTr 78
 |||:::|||||
DB 653 CCACCTCATGTTACAGGAATCCGCCGAGACCACTTATATATATATCAAGAAAGAGCA 594
 |||
QY 78 pProValAlaLeuTyrIleThrProValSerSerAlaGlyClnLeuValIleLysAlaG 98
 |||:::|||||
DB 593 AAGCGGTGCTGCTTCCAGACTCTTCAGATGTTCTGCAGCCCAAGTCTGTATATATCAAA 534
 |||
QY 98 Y-----SerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyr 112
 |||:::|||||
DB 533 TTATATCCCAAGCCACAAAGAACTCTTATGATACACTGCTTCAAGGACTCATTAAGCT 474
 |||
QY 112 rAsnSerAspAsp-----PheGln-PheValTyrPasnIleTyrAlaAsnAsnAspVal 130
 |||:::|||||
DB 473 TCACAGCTCACACAGATTTCCTCAAAATTTGTATGCATATTC----- 433
 |||
QY 130 aValProThrGlnGlyCysAspValSerAlaArgAspValThrValThrLeuProAspT 150
 |||
DB 432 -----TAAAGTCCGCTTACATATCCCACTGT 408
 |||
QY 150 YrArgGlySerValProIlePro-----LeuThrValTyr 161
 |||:::|||||
DB 407 AC-----ATCATATCCCATCCACGCTTTGTCTGACACTGTAT 370
 |||

RESULT 11

LOCUS BG621049 834 bp mRNA linear EST 18-APR-2001
DEFINITION 602616925F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4730406 5',
 mRNA sequence.
ACCESSION BG621049
VERSION BG621049.1 GI:13672420
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
 Tissue Procurement: CLOWNTECH Laboratories, Inc.
 CDNA Library Preparation: CLOWNTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.


```

Db      539 ATTCTAATATGCTGTACACTTGAATAAT-----ATTCAATTAACA 580
QY      160 Val 160
Db      581 TTA 583

RESULT 13
BO970561
LOCUS    BO970561
DEFINITION OHB42H01.yg.ab1 OH-ABCDI sunflower RHAB01 Helianthus annuus cDNA
ACCESSION BO970561
VERSION   BO970561.1 GI:22388082
KEYWORDS  EST.
SOURCE    Common sunflower.
ORGANISM  Helianthus annuus.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;
Heliantheae; Helianthus.
1 (bases 1 to 756)
Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,D., Ellison
,P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
Lettuce and sunflower ESTs from the Composite Genome Project
http://comphenomics.ucdavis.edu/
Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig OH-CA-Contig1697, see http://cgpsdb.ucdavis.edu/
for details.
Plate: OHB42 row: H column: 01.
Location/Qualifiers
1..756
/organism="Helianthus annuus"
/cultivar="RHA801"
/db_xref="taxon:4232"
/clone="OHB42H01"
/clone_1lb="OH-ABCDI sunflower RHA801"
/lab_host="E. coli"
/note="Vector: pBRCONASf1AB. The library was constructed
from 11 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpsdb.ucdavis.edu/
TAG_SEQ=Not found"
BASE COUNT 214 a 148 c 172 g 222 t
ORIGIN
Alignment Scores:
Pred. No.: 18 4 Length: 756
Score: 83.00 Matches: 35
Percent Similarity: 43.48% Conservative: 25
Best Local Similarity: 25.36% Mismatches: 56
Query Match: 9.79% Indels: 22
DB: 14 Gaps: 6
US-09-900-575-29_COPY_26_186 (1-161) x BO970561 (1-756)
QY 30 TyrValThrLeuGlnArgGlySerAlaTyrGlyValLeuSerAspSerGlyThr 49
Db 40 TGTACTGCGATTAAACGAGCTTCTGTACATTATCTGATGTCATCTCTACTCGCTTAT 99

```

```

QY 50 ValIysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgValVal 69
Db 100 GGAATGTACACG---CTCAATATATGAATG---CTCAGTAAAGAGCTCCCAAGTGC--- 150
QY 70 TyrAsnSerArgThrAspLysProTyrProValAlaLeuTyrLeuThrProValSerSer 89
Db 151 -----AACCGAACCGAACGACCATGTTAATAGGTGATTAGCATTTGCTCATTTGATAGT 204
QY 90 AlAGlGlyLeuValIleIleLysAlaGlySerLeuIleAlaValIleLeuArgGlnThr 109
Db 205 AGCTATGACATCATATATATGAGAGTGAACCATGAGAGTATGACGAACAATATTTT 264
QY 110 AsnAsnTyrAsnSerAsp---AspPheGlnPheValTyrAsnIleTyrAlaAsnAsp 128
Db 265 GTTCAGTACAAAGTTGATGTTGTCTCTGACACGTTATGCTTACGACGACGAC 324
QY 129 ValVal-----ValProThrGlyGlyCysAspValSerAlaArgAsp 142
Db 325 CGGGATCGAATATATAGCTTACAAATATCGAAACGGGTGTGCAAT----- 369
QY 143 ValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
Db 370 -----CCTAATATATGATGATCTGCTCCCTGTTTACATTAACCGTT 408

RESULT 14
BF557712
LOCUS    BF557712
DEFINITION UI-R-C0-hb-d-06-0-UI.r1 UI-R-C0 Rattus norvegicus cDNA clone
ACCESSION BF557712
VERSION   BF557712.1 GI:11667442
KEYWORDS  EST.
SOURCE    Norway rat.
ORGANISM  Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 413)
Bonaldi,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
cDNA Library Preparation: M.B. Soares lab clone distribution:
clones will be available through Research Genetics (www.resgen.com)
this clone is also available through the I.M.A.G.E. Consortium at
LNL (info@image.llnl.gov). IMAGE ID- 1773173
Seq primer: M13 Forward.
Location/Qualifiers
1..413
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C0-hb-d-06-0-UI"
/clone_1lb="UI-R-C0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; The UI-R-C0
library is a subtracted library derived from the UI-R-A1
and UI-R-E1 libraries. The UI-R-A1 library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, and muscle. The UI-R-E1
library consisted of a mixture of individually tagged

```

normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dr track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-CO) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-Al and UI-R-EI clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-Al and UI-R-EI library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-CO library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996).

BASE COUNT 104 a 129 c 82 g 98 t
ORIGIN

Alignment Scores:

Pred. No.: 8.28 Length: 413
Score: 82.50 Matches: 28
Percent Similarity: 37.30% Conservative: 19
Best Local Similarity: 22.22% Mismatches: 42
Query Match: 9.73% Indels: 37
DB: 12 Gaps: 3

US-09-900-575-29_COPY_26_186 (1-161) x BF557712 (1-413)

QY 13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
DB 40 TTACAACTGCTGTCTCTGTGCAAAACGATGCCCGGAAGTTAGAC----- 87
QY 33 LeuGlnArgGlySerAlaTyrGlyValLeuSerAsnPheserGlyThrValIlystyr 52
DB 87 ----- 87
QY 53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg-----Val 68
DB 88 -----TATCACCACACCTCCACACCCCGCAAGAGACACACCCACTCAA 138
QY 69 ValTyrAsnSerArgThrAspLysProTyrProValAlaLeuTyrLeuThrProValSer 88
DB 139 GTCTCTCTCTCCACAGTGAACCGTGTGTACTAGTACGATTGTGACACCAACAGTT 198
QY 89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
DB 199 CTAACCTGTGGACGATTTGATATCTCCACACTAGTATTAGCACTCACTCAGACCGA 258
QY 109 ThrAsnAsnTyrAsnSerAspPheGlnPheValTyrPasnIleTyrAlaAsnAsp 128
DB 259 GAATCAAAATTCGTCTGTCAGTAGACGTCTCTT-----CAGGATGAC 300
QY 129 ValValValProThrGly 134
DB 301 ACCATTGTCCCATAGGA 318
RESULT 15
BG664251
LOCUS BG664251
DEFINITION DRABBC11 Rat DRG Library Rattus norvegicus cDNA clone DRABBC11 5',
ACCESSION BG664251
VERSION BG664251.1 GI:13886173
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Rattus norvegicus; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 501)
AUTHORS Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.

TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome
Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
Pudong New Area, P.R.China. Please contact with Zhang Xu
(xu.zhang@ion.ac.cn) or Han Zeyuang (hanzheyu@ion.sh.cn)
PCR Primers
FORWARD: T7
BACKWARD: T7
Seq primer: T3
POLYA-No.

FEATURES
source
Location/Qualifiers

1..501
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/note="Total RNA was isolated from hypothalamus and transcribed into cDNA, which was then used as template in PCR. The PCR products were separated on sequencing gel. The differential bands were cut, reamplified, cloned into pMD18-T vector and confirmed by Northern blot."

BASE COUNT 130 a 147 c 109 g 114 t 1 others

Alignment Scores:

Pred. No.: 11.2 Length: 501
Score: 82.50 Matches: 28
Percent Similarity: 37.30% Conservative: 19
Best Local Similarity: 22.22% Mismatches: 42
Query Match: 9.73% Indels: 37
DB: 12 Gaps: 3

US-09-900-575-29_COPY_26_186 (1-161) x BG664251 (1-501)

QY 13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
DB 189 TTACAACTGCTGTCTCTGTGCAAAACGATGCCCGGAAGTTAGAC----- 236
QY 33 LeuGlnArgGlySerAlaTyrGlyValLeuSerAsnPheserGlyThrValIlystyr 52
DB 236 ----- 236
QY 53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg-----Val 68
DB 237 -----TATCACCACACCTCCACACCCCGCAAGAGACACCCACTCAA 287
QY 69 ValTyrAsnSerArgThrAspLysProTyrProValAlaLeuTyrLeuThrProValSer 88
DB 288 GTCTCTCTCTCCACAGTGAACCGTGTGTACTAGTACGATTGTGACACCAACAGTT 347
QY 89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
DB 348 CTAACCTGTGGACGATTTGATATCTCCACACTAGTATTAGCACTCACTCAGACCGA 407
QY 109 ThrAsnAsnTyrAsnSerAspPheGlnPheValTyrPasnIleTyrAlaAsnAsp 128
DB 408 GAATCAAAATTCGTCTGTCAGTAGACGTCTCTT-----CAGGATGAC 449

QY 129 ValValProthrgly 134
 DB 450 ACCATTGTCCTCATAGCA 467

RESULT 16
 LOCUS B6664658 518 bp mRNA linear EST 30-APR-2001
 DEFINITION DRABGH01 Rat DRG library Rattus norvegicus cDNA clone DRABGH01 5',
 mRNA sequence.
 ACCESSION B6664658
 VERSION B6664658.1 GI:13866580
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE
 1 (bases 1 to 518)
 Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.,
 Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang
 X., Chen,Z., Han,Z.G. and Zhang,X.
 Identification of gene expression profile of dorsal root ganglion
 in the rat peripheral exotomy model of neuropathic pain
 Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)

TITLE 2205133
 JOURNAL Contact: Zhang Xu
 MEDLINE Laboratory of Sensory System
 COMMENT Institute of Neuroscience
 320 Yue Yang Road, Shanghai 200031, P.R.China
 Tel: 86-21-64748700-121
 Fax: 86-21-64713446
 Email: xu.zhang@lon.ac.cn
 This clone is also available at Chinese National Human Genome
 Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
 Pudong New Area, P.R.China. Please contact with Zhang Xu
 (xu.zhang@lon.ac.cn) or Han Zeguang (hanzegu@nc.sh.cn)
 PCR Primers
 FORWARD: T3
 BACKWARD: T7
 Seq primer: T7
 POLYA-No.

FEATURES
 source Location/Qualifiers
 1..518
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="DRABGH01"
 /clone_1lb="Rat DRG Library"
 /sex="male"
 /tissue_type="dorsal root ganglion"
 /dev_stage="adult"
 /note="Total RNA was isolated from hypothalamus and
 transcribed into cDNA, which was then used as template in
 PCR. The PCR products were separated on sequencing gel.
 The differential bands were cut, reamplified, cloned into
 pMD18-T vector and confirmed by Northern blot."

BASE COUNT 132 a 150 c 112 g 124 t

ORIGIN
 Alignment Scores:
 Pred. No.: 11.8 Length: 518
 Score: 82.50 Matches: 28
 Percent Similarity: 37.30% Conserved: 19
 Best Local Similarity: 22.22% Mismatches: 42
 Query Match: 9.73% Gaps: 37
 DB: 12 Gaps: 3

US-09-900-575-29_COPY_26_186 (1-161) x B6664658 (1-518)

QY 13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleMetArgTyrValThr 32
 DB 189 TTACACACTGCTTCTCTGTGCAACGATGCCCGGAGAGTTAGACC----- 236

QY 33 LeuGlnArgGlySerAlaTyrGlyValLeuSerAsnPheSerGlyThrValIstYr 52
 DB 236 ----- 236

QY 53 SerGlySerSerTyrProPheProThrSerGluThrProArg-----Val 68
 DB 237 -----TATACACCCACACCCCTCCACACACCCAGCCAAAGACACCCCACTCA 287

QY 69 ValTyrAsnSerArgThrAspTyrProTyrProValAlaLeuTyrLeuThrProValSer 88
 DB 288 GTCTCTTCTTCCACAGTACCGGTGTTTACTACAGTACGTATTTGACACCAACACT 347

QY 89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValIleLeuArgGln 108
 DB 348 CTAACGTGGTGAACGATTAAGATTCGACACACTAAGTATTAAGTCTTAACACACCA 407

QY 109 ThrAsnAspTyrAsnSerAspAspPheGlnPheValTyrAsnIleTyrAlaAsnAsp 128
 DB 408 GAATCAAAATTCCTGCTCAGTAGACGCTCTCTT-----CAGATGAC 449

QY 129 ValValProthrgly 134
 DB 450 ACCATTGTCCTCATAGCA 467

RESULT 17
 LOCUS BF072843/c 530 bp mRNA linear EST 25-APR-2001
 DEFINITION 219330 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION BF072843
 VERSION BF072843.1 GI:10866354
 KEYWORDS EST.

SOURCE cow.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovine; Bos.

REFERENCE
 1 (bases 1 to 530)
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
 G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McCorm,C.G.,
 Pette,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
 Keefe,J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)

TITLE 21180013
 JOURNAL Contact: Smith TPL
 MEDLINE USDA, ARS, US Meat Animal Research Center
 COMMENT PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -m1nscore 18
 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGCAACAGCTATGACCA
 BACKWARD: GTTTCACAGTCACGAGC
 Plates: 77 row: B column: 2
 Seq primer: ATTAGTGACACTATAG.

FEATURES
 source Location/Qualifiers
 1..530
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_1lb="MARC 2BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: PCMV SPOR6; site.1: NotI; site.2: SalI;
 library made from pooled tissue from testis, thymus,
 semitendinosus muscle, longissimus muscle, pancreas,
 adrenal, and endometrium."

BASE COUNT 126 a 131 c 182 g 91 t

REFERENCE 1 (bases 1 to 696)
AUTHORS Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
TITLE Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
MEDLINE 22056133
COMMENT Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzegu@ion.sh.cn)
PCR Primers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA-No.

FEATURES
source Location/Qualifiers
1..696
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRACE11"
/clone_lib="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/note="Total RNA was isolated from hypothalamus and transcribed into cDNA, which was then used as template in PCR. The PCR products were separated on sequencing gel. The differential bands were cut, reamplified, cloned into PMD18-T vector and confirmed by Northern blot."

BASE COUNT 168 a 187 c 162 g 174 t 5 others
ORIGIN

Alignment Scores:
Pred. No.: 18.6 Length: 696
Score: 82.50 Matches: 28
Percent Similarity: 37.30% Conservative: 19
Best Local Similarity: 22.22% Mismatches: 42
Query Match: 9.73% Indels: 37
DB: 12 Gaps: 3

US-09-900-575-29_COPY_26_186 (1-161) x BG665535 (1-696)

QY 13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
111
120 TTACACCTGCTTCTCTCTGTGCACAAAGATGCCCGGAAAGTTAGACC----- 167
33 LeuGlnArgIleSerAlaTyrGlyValLeuSerAsnPheSerGlyThrValIleTyr 52
167 ----- 167

QY 53 SerGlySerTyrProPheProThrThrSerGluThrProArg-----Val 68
111
168 -----TATCACCCACACCTCCACACCCACCCCAAAAGACAGACCCCACTCA 218
69 ValIyrAsnSerArgThrAspIleProTyrProValAlaLeuTyrIleThrProValSer 88
111
219 GTCTCTTCTTCCACAGTCAACGGGGTTGTACTAGTACGATTTTGACACCAAGCTT 278
89 SerIleGlyIleValIleAlaGlySerLeuIleAlaValIleLeuArgGln 108
111
279 CTAACGTGTGACAGCATTTAGATTCGCACACTAGTATTACATCACTCAACACGACGA 338
109 ThrAsnAspTyrAsnSerAspAspPheGlnPheValIleTyrAsnIleTyrAlaAsnAsp 128

Db 339 GAATCAAAATTCCTGCTCAGTACAGACGTCCTCTT-----CAGATGAC 380
QY 129 ValValValProThrGly 134
Db 381 ACCATGTGCCCCATATGAGA 398

RESULT 22
BG669393
LOCUS
DEFINITION DNAAF07 Rat DRG library Rattus norvegicus cDNA clone DNAAF07 5', mRNA sequence.
ACCESSION BG669393
VERSION BG669393.1 GI:13891315
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE 1 (bases 1 to 697)
AUTHORS Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
TITLE Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
MEDLINE 22056133
COMMENT Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzegu@ion.sh.cn)
PCR Primers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA-No.

FEATURES
source Location/Qualifiers
1..697
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRNAF07"
/clone_lib="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/note="Total RNA was isolated from hypothalamus and transcribed into cDNA, which was then used as template in PCR. The PCR products were separated on sequencing gel. The differential bands were cut, reamplified, cloned into PMD18-T vector and confirmed by Northern blot."

BASE COUNT 173 a 189 c 159 g 169 t 7 others
ORIGIN

Alignment Scores:
Pred. No.: 18.7 Length: 697
Score: 82.50 Matches: 28
Percent Similarity: 37.30% Conservative: 19
Best Local Similarity: 22.22% Mismatches: 42
Query Match: 9.73% Indels: 37
DB: 12 Gaps: 3

US-09-900-575-29_COPY_26_186 (1-161) x BG669393 (1-697)

QY 13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32


```

Query Match: 9.73% Indels: 37
DB: 12 Gaps: 3
us-09-900-575-29_copy_26_186 (1-161) x BG673676 (1-747)

QY 13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
    |||  ::  |||  |||||  |||||  ::  |||
DB 109 TTACAACTGCTGTCTCTGTGCMAACGATGCCCGGAAGTTAGACC----- 156
QY 33 LeuGlnArgGlySerAlaIleGlyValIleuSerAsnSereGlyThrValIleTyr 52
156 ----- 156
QY 53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg-----Val 68
    ::|||  |||  |||||  ::|||  ::|||  ::|||
DB 157 -----TATCACCACCAACCTCCACACCCAGCCCAAGAGACAGCCCACTCAA 207
QY 69 ValITyrAsnSerArgThrAspLysProTrrProValAlaIleuTyrLeuThrProValSer 88
    |||  ::|||  ::|||  ::|||  ::|||  ::|||
DB 208 GTCTCTTCTTCCACAGTGAACCGTGTGTACTAAGACGATTTTGACACACAGCTT 267
QY 89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
    |||||  ::|||  ::|||  ::|||  ::|||  ::|||
DB 268 CTAAGTGGTGGACGATTAAGATCTGCACACTAAGTATTAGATCTTAAGTCAACGACGCA 327
QY 109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValITrPasnIleTyrAlaAsnAsnAsp 128
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
DB 328 GAATCAAAATTTCTGCTGCTGACGTAGACGTCCTCTT-----CAGATGAC 369
QY 129 ValValIleProThrGly 134
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
DB 370 ACCATTGTCCCATAGGA 387

RESULT 26
BG671529 789 bp mRNA linear EST 30-APR-2001
LOCUS DNRBT006 Rat DRG Library Rattus norvegicus cDNA clone DNRBT006 5',
DEFINITION mRNA sequence.
ACCESSION BG671529
VERSION BG671529.1 GI:13893628
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 789)
AUTHORS Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.,
Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang
,X., Chen,Z., Han,Z.G. and Zhang,X.
Identification of gene expression profile of dorsal root ganglion
in the rat peripheral axotomy model of neuropathic pain
Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
22056133
COMMENT Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome
Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
Pudong New Area, P.R.China. Please contact with Zhang Xu
(xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
PCR Primers
FORWARD: T3
BACKWARD: T7
SEG PRIMER: T3
POLYA-NO.

FEATURES
Source 1..789 Location/Qualifiers
/organism="Rattus norvegicus"

/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DNRBT006"
/clone_id="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/note="Total RNA was isolated from hypothalamus and
transcribed into cDNA, which was then used as template in
PCR. The PCR products were separated on sequencing gel.
The differential bands were cut, reamplified, cloned into
pMD18-T vector and confirmed by Northern blot."
BASE COUNT 193 a 213 c 192 g 191 t
ORIGIN

Alignment Scores:
Pred. No.: 22.6 length: 789
Score: 82.50 Matches: 28
Percent Similarity: 37.30% Conservative: 19
Best Local Similarity: 22.22% Mismatches: 42
Query Match: 9.73% Indels: 37
Gaps: 3

us-09-900-575-29_copy_26_186 (1-161) x BG671529 (1-789)
QY 13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
    |||  ::  |||  |||||  |||||  ::  |||
DB 164 TTACAACTGCTGTCTCTGTGCMAACGATGCCCGGAAGTTAGACC----- 211
QY 33 LeuGlnArgGlySerAlaIleGlyValIleuSerAsnSereGlyThrValIleTyr 52
211 ----- 211
QY 53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg-----Val 68
    ::|||  |||  |||||  ::|||  ::|||  ::|||
DB 212 -----TATCACCACCAACCTCCACACCCAGCCCAAGAGACAGCCCACTCAA 262
QY 69 ValITyrAsnSerArgThrAspLysProTrrProValAlaIleuTyrLeuThrProValSer 88
    |||  ::|||  ::|||  ::|||  ::|||  ::|||
DB 263 GTCTCTTCTTCCACAGTGAACCGTGTGTACTAAGTACGATTTTGACACACAGCTT 322
QY 89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
    |||||  ::|||  ::|||  ::|||  ::|||  ::|||
DB 323 CTAAGTGGTGGACGATTAAGATCTGCACACTAAGTATTAGATCTTAAGTCAACGACGCA 382
QY 109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValITrPasnIleTyrAlaAsnAsnAsp 128
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
DB 383 GAATCAAAATTTCTGCTGCTGACGTAGACGTCCTCTT-----CAGATGAC 424
QY 129 ValValIleProThrGly 134
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
DB 425 ACCATTGTCCCATAGGA 442

RESULT 27
BI079789 1312 bp mRNA linear EST 20-JUN-2001
LOCUS 602876103F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5007745 5',
DEFINITION mRNA sequence.
ACCESSION BI079789
VERSION BI079789.1 GI:14498119
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1312)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rtmail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.

```


Unpublished (2000)

Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113

Location/Qualifiers
1. .597

```
1. .597
/organism="Clona intestinalis"
/db_xref="taxon:7719"
/clone="cliv40d02"
CloneLib="Norl Satoh unpublished cDNA library, larva"
```

```

/note="Vector: pBluescript SK"
BASE COUNT      149 a      184 c      117 g      147 t

```

Prod. No.:	16.8	Length:	3
Score:	82.00	Matches:	5
Percent Similarity:	40.31%	Conservative:	1
Best Local Similarity:	27.91%	Mismatches:	6
Query Match:	9.67%	Indels:	1
Ds:			

OS-09-900-5/5-29_COPY_26_186 (1-161) x AV984947 (1-597)

ValLeuSerAlaHisPheSerGlyIleValLysTyrSerGlySerSerTyr 57

[illegible]

255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766

Ov 73 ArqThrasnlvsPcTrrPrvaJal at Outt r oumbdrevyJ deGant) - d) uat; ss

Db 295 AAGCTGATCCACCTGGTACGTGGCGGACGAAGAAACGCAATGCTGTACACCCGCC 335

QY 93 LeuVal-----IleLysAlaGlySerLeuIleAlaValLeuIleLeuArg 107

Db 235 TTGGTCAAGTACTACATTAACTACGGGTGGTGGTTAGTAGTCACTCGACAAG-CGT 177

108 GlnThrAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsn 127

22 170 GCGAAGGGAACCTGTGGGACCACTGGTGGCAATGTGACTGGAACTCTTGGCAGATACA 117

[illegible][illegible][illegible]

RESULT 31

AV996923/C	LOCUS	AV996923	614 bp	mpna	110000	ECM 15-MAR

Intestinalis cDNA clone c11v29p03 5' mRNA sequence

VERSION AV996923.1 GI:19488257

SOURCE	Ciona intestinalis.
ORGANISM	Ciona intestinalis.

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogastralia; Phlebobranchia; Clonidae; Clon

AUTHORS Satoh, N., Satou, Y., Kohara, Y., and Shin-I, T.

0
 1
 2
 3
 4
 5
 6
 7
 8
 9
 A
 B
 C
 D
 E
 F
 G
 H
 I
 J
 K
 L
 M
 N
 O
 P
 Q
 R
 S
 T
 U
 V
 W
 X
 Y
 Z

JOURNAL
COMMENT

Unpublished (2000)
Contact: Nori Satoh
Department of zoology
Kyoto University
Sakyo-u, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-703-1113
Email: satoh@scs.dan.zoo1.kyoto-u.ac.jp
Location/Qualifiers

BASE COUNT	185 a	215 c	132 g	154 t	3 others
ORIGIN					

Alignment Scores:	
Pred. No.:	21
Score:	82.00
Percent Similarity:	40.31%
Best Local Similarity:	27.91%
Query Match:	9.67%
DB:	10
Length:	689
Matches:	36
Conservative:	16
Mismatches:	61
Indels:	17
Gaps:	4

US-09-900-575-29_COPY_26_186 (1-161) x AV970876 (1-689)

```

Oy      38 AlaTyrGlyGlyValLeuSerAsnSphHeeRgYThrValLysTYrSerLYrSerSerTYr    57
Db      580 GGTGATGGGGGGCGGCACATAATCCGATCATGTAAATTGGTTTCACAATTGACCCTTG    521
Oy      58 ProPheProThrThrSerGIuThrProArgValValTYrAsnSer-----72
Db      520 GACTGCCAATTTGTCAAAATGCTCTCCGGCAACGATATATCCAGCTTGGCATTGGATGTA    461
Oy      73 ArgThrAspLysProTrpProValAlaIleuTYrLeuThrProValSerSerAlaGlyLy    92
Db      460 AAGCTGATCCACCACTGCTAGCTGGGGCGAGAAGAACCCCCAAATGCTGTGAACAGGGGGG    401
Oy      93 LeuVal-----lleySalAglySerlleAlaValLeuileuArg    107
Db      400 TTGTGTCGAAGTACTACATTAACTACACGGCTGGTGGTAGTGATGACTCGAACAAG-CGT    342
Oy      108 GluThrAsnAsnTYrAsnSerAspSphEgInpheValITrpAsnIlleTYrAlaAsnAsn    127
Db      341 GCCGAGGGAACACTGTTGGGCAACACTGGCGCATGACTGGAACTCTGGCAGAAATACA    282
Oy      128 AspValValValProThrglyglyCyaspavalSer-----AlaArgspvalThr    144
Db      281 GACCAAGATr-----GGTTGGTGTGAATGTGCTCCGTGGCACATGACTTGTGG    231
Oy      145 ValThrLeuProAspTYrArgGlySer    153
Db      230 ATTTCCTGGATGACGTAAAGGCGCTTG    204

```

RESULT 36	AL667637/c	939 bp	mRNA	linear	EST 11-JAN-2002
LOCUS	AL667637/c				
DEFINITION	AL667637	directional larval cDNA	library	Ciona intestinalis	cDNA
ACCESSION	AL667637	clone 022ZC11 5',			mRNA sequence.
VERSION	AL667637				
KEYWORDS	AL667637.1	GI:18134544			
SOURCE	EST.				
ORGANISM	Ciona intestinalis				
	Ciona intestinalis				
	Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;				
	Phlebobranchia; Clonidae; Ciona.				
	1 (bases 1 to 939)				
REFERENCE	Genoscope.				
AUTHORS	Ciona intestinalis	directional larval cDNA	library		
TITLE					

JOURNAL
COMMENT

Unpublished (2002)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segruff@genoscope.cns.fr, Web : www.genoscope.cns.fr
IMPORTANT: this sequence may contain errors. The Clona intestinalis
library from which the clone was isolated may be contaminated with
cDNAs from bacteria or other Eukaryote.
Directional larval cDNA library originate from Dr. M. Branno,
Stazione A. Dohrn, Naples, Italy, and was prepared in
pBluescript2SK⁺. Coordinates: Genbank: AF151518

BASE COUNT	247 a	279 c	187 g	225 t	1 others
ORIGIN					

Alignment Scores:	
Pred. No.:	34
Score:	82.00
Percent Similarity:	40.31%
Best Local Similarity:	27.91%
Query Match:	9.67%
DB:	
	9
	Gaps:
	4
	Length:
	939
	Matches:
	36
	Conservative:
	16
	Mismatches:
	61
	Indels:
	17
	Gaps:
	4

US-09-900-575-29_COPY_26_186 (1-161) x AL667637 (1-939)

```

Oy      38 AlaTyrGlyGlyValLeuSerAsnPheserGlyThrValLysTyrSerLysSerTyr 57
      ||| ||||| ||||| ||||| |||||
Db      771 GCGTANGCGCGCGCGACATCTAATAGCATGATGTAATTTTGCTTCAAAATTTGACCGTTG 712
Oy      58 ProPheProThrThrSerGluThrProArgValValTyrAsnSer-----72
      ||| ||||| ||||| ||||| |||||
Db      711 GACTGCGACAGTTGTGAAAGTGTCTCCGGCAACGATATATACAGGTGGCATTTGATGTA 652
Oy      73 ArgTyrAspLysProThrProValAlaLeuTyrLeuThrProValSerSerAlaGly 92
      ||| ||||| ||||| ||||| ||||| ||||| |||||
Db      651 AAGCGTGAATCAACACTGTGACGCGGGCGGAGCAAGAACCCCAATGCTGTGACAGGGGG 592
Oy      93 LeuVal-----LleuSalagLysSerLeuIleAlaValLeuLeuArg 107
      ||||| ||||| ||||| ||||| ||||| |||||
Db      591 TTGTGTGCACACTACTACTATTAACTACGCGGTGGTGGTTAGTGAATCACTGCAGCAAG-CGT 533
Oy      108 GluThrAsnAsnTyrAsnSerAspPheGlnPheValTyrAsnIleTyrAlaAsnAsn 127
      ||| ||||| ||||| ||||| ||||| ||||| |||||
Db      532 GCGGAGGGAACACTGTGGGCAACACTGCTGGCACTGACCTGCAGCACTCTTGCGAGATACA 473
Oy      128 AspValValValProThrGlyGlyCysAspValSer-----AlaArgAspValThr 144
      ||| ||| ||| ||||| ||||| ||||| ||||| |||||
Db      472 GACAAAGATT-----GGTTGGTGTGAATTGTCCAGTTCCTTGCGACAGATCACTTGTGG 422
Oy      145 ValThrLeuProAspTyrArgGlySer 153
      ||||| ||||| ||||| ||||| |||||
Db      421 ATTTCCTTGGATGACGTAAAGGCCCTCG 395

```

RESULT	37
LOCUS	BG666005
DEFINITION	BG666005 544 bp mRNA linear EST 30-APR-2001 DRACMA09 Rat DRG Library Rattus norvegicus CDNA clone DRACMA09_5'
ACCESSION	BG666005
VERSION	BG666005.1 GI:13887927
KEYWORDS	EST.
SOURCE	Norway rat. <i>Rattus norvegicus</i>
ORGANISM	Eukaryota; Metazoa; Chordata; Cranialia; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; <i>Rattus</i> .
REFERENCE	1 (bases 1 to 544)

AUTHORS
Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.,
Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang
,X., Chen Z., Han,Z.G. and Zhang,X.
TITLE
Identification of gene expression profile of dorsal root ganglion
in the rat peripheral axotomy model of neuropathic pain
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 99 (112), 8360-8366 (2002)
MEDLINE
22056133
COMMENT
Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome
Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
Pudong New Area, P.R.China. Please contact with Zhang Xu
(xu.zhang@ion.ac.cn) or Han Zeguang (hanzeguang.sh.cn)
PCR Primers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA-No.

FEATURES
source
1. 544
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRACMA09"
/clone_lib="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/note="Total RNA was isolated from hypothalamus and
transcribed into cDNA, which was then used as template in
PCR. The PCR products were separated on sequencing gel.
The differential bands were cut, reamplified, cloned into
pMD18-T vector and confirmed by Northern blot."

BASE COUNT
144 a 147 c 124 g 129 t
ORIGIN

Alignment Scores:
Pred. No.: 16.7 Length: 544
Score: 81.50 Matches: 28
Percent Similarity: 38.10% Conservative: 20
Best Local Similarity: 22.22% Mismatches: 41
Query Match: 9.61% Indels: 37
Gaps: 4

US-09-900-575-29_COPY_26_186 (1-161) x BG666005 (1-544)

QY 13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
Db 225 TTACAACTGCTGTGCTGTTCAACGATGCCCG----- 260

QY 33 LeuGlnArgGlySerAlaIleTyrGlyValLeuSerAsnPheSerGlyThrValLeuTyr 52
Db 261 ----- 275

QY 53 SerGlySerTyrProPheProThrThrSerGluThrProArg-----Val 68
Db 276 -----CACCCACACCCCTGCCACACACCCGCCAAGAGACAGCCCACTCAA 323

QY 69 ValTyrAsnSerArgThrAspLysProTyrProValAlaLeuTyrLeuThrProValSer 88
Db 324 GTCTCTTCTTCACAGTGAACCGGTTGTACTAGTACGATATTGACACCAACAGT 383

QY 89 SerAlaGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
Db 384 CTAGCTGTGGAACGATTACATCCGACACTAGATATTGACTTCACTCAACCGA 443

QY 109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTyrAsnIleTyrAlaAsnAsp 128
Db 109 ----- 128

Db 444 GAACTAAATTCCTGCTCAGTAGAC-----GCTACTTCAGAGATGAC 485
QY 129 ValValAlaProThrGly 134
Db 486 ACCATTGTCCTCATAGGA 503

RESULT 38
AA969177/c 443 bp mRNA 11near EST 07-JUL-1998
LOCUS
DEFINITION
op58c03.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1581028.3 Similar to gb:U02570.1111 ALU CLASS C WARNING
ENTRY 1111 (HUMAN); mRNA sequence.

ACCESSION
AA969177
AA969177.1 GI:3144357
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 443)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
Tumor Gene Index
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
CONTACT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 603 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.

FEATURES
source
1. 443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1581028"
/clone_lib="Soares_NFL_T_GBC_S1"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker. Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbH129, testis NHT, and B-cell
NCI-CCAP-GC81) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT
122 a 83 c 136 g 102 t
ORIGIN

Alignment Scores:
Pred. No.: 14 Length: 443
Score: 81.00 Matches: 28
Percent Similarity: 43.75% Conservative: 14
Best Local Similarity: 29.17% Mismatches: 44
Query Match: 9.55% Indels: 10
Gaps: 4

US-09-900-575-29_COPY_26_186 (1-161) x AA969177 (1-443)

QY 64 GluThrProArgValValTyrAsnSerArgThrAspLysProTyrProValAlaLeuTyr 83
Db 328 GACTCTCCCAACACCATC-----AACATACATCCAAACCCCTGAGCCCTGCTGCTG 275

QY 84 LeuThrProValSerSerAlaGly-----GlyLeuValIleLysAlaGly 98
Db 274 CTGATCTTCATCACTGCTTGTGCTAGCCTGCGGACGGGCTGCTGCTCCGCC 215

QY 99 SerLeuIleAlaValLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGln 118
Db 99 ----- 118

```
Db 214 TCCTTTGCTGCACACTACGCTC-----ACCGACCATTCATMGAGACCTTCCGA 161
Qy 119 PheValTrpAsnIleTyrAlaAsnAspValValAlaProThrGlyGlyCysAspVal 138
Db 160 TCTATTGTGCTCTTCATGTCCTCCCACTGGTCTGTACCCAGAGAGGCGTCTTGAC 101
Qy 139 SerAlaAspAspVal---ThrValThrLeuProAspTyrArgGlySer 153
Db 100 TGTGTAATTCACAGTTCACATTCACACTTAATGACTTCCTTGCGACC 53

RESULT 39
BH519770 594 bp DNA linear GSS 13-DEC-2001
BH519770
LOCUS BOGDG45TF BOGD Brassica oleracea genomic clone BOGDG45, DNA
DEFINITION
sequence.
ACCESSION BH519770
VERSION BH519770.1 GI:17727855
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea.
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLE Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
JOURNAL 1 (bases 1 to 594)
COMMENT Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSs: BOGDG45TR
Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: ctown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
FEATURES
source 1..594
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGDG45"
/clone_lib="BOGD"
/Note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 125 a 174 c 144 g 151 t
ORIGIN
Alignment Scores: 22 Length: 594
Pred. No.: 25 Matches: 25
Score: 81.00 Conservative: 16
Percent Similarity: 46.07% Mismatches: 34
Best Local Similarity: 28.09% Indels: 14
Query Match: 9.55% Gaps: 2
DB: 17

US-09-900-575-29_COPY_26_186 (1-161) x BH519770 (1-594)
Qy 24 ProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyValLeu 43
Db 305 CCGTCCACGGCTACCTCTCCCGCTTCACAGATTCCTCGACGTCGCGGTTGATC 364
Qy 44 SerAsnPheserGlyValIysTyrSerGlySerSerIyrPrope-----Pro 60
Db 365 TCGAATTTGAACCGATCGTCCCTTCACGTCATCGATTGACCGTACGGGTTCATACCT 424
Qy 61 ThrThrSerGluThrProAspValValIleTyrAsn-SerArgThr-AspIysProThrPro 80
Db 425 AATACGATTCGATGAGAGAAAGCCAGCATTTGGTAAAGCGTCTCGCATG----- 480
Qy 80 alaIaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleIysAlaGlySerL 100
```

```
Db 481 -----ACCTCTCGGAGGCTTGGGATTCAGAGAGAGCTG 517
Qy 100 eulIeAlaValIleuIleuArg 107
Db 518 TGTTCGCCGCTGCTCTTGAGA 540

RESULT 40
AG072173 669 bp DNA linear GSS 03-NOV-2001
AG072173
LOCUS Pan troglodytes DNA, clone: PTB-063J11.F, genomic survey sequence.
DEFINITION
ACCESSION AG072173
VERSION AG072173.1 GI:16623975
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC library clone:PTB-063J11.F.
ORGANISM Pan troglodytes
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
AUTHORS Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 669)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
REFERENCE and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
AUTHORS 1-7-22 Suehiro-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
TITLE (E-mail:chimbesc@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
JOURNAL Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pRS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
FEATURES
source 1..669
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-063J11.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

BASE COUNT 155 a 166 c 135 g 213 t
ORIGIN
Alignment Scores: 26.5 Length: 669
Pred. No.: 42 Matches: 42
Score: 81.00 Conservative: 24
Percent Similarity: 37.93% Mismatches: 55
Best Local Similarity: 24.14% Indels: 53
Query Match: 9.55% Gaps: 10
DB: 17

US-09-900-575-29_COPY_26_186 (1-161) x AG072173 (1-669)
Qy 2 ValValAsnValGlyIlnAsnLeu-----ValValAspLeuSer 14
Db 145 GTTAAAGAACCAAGCAGATGTGACACTTGAATCTGACCTGATCTGATCTCAAT 204
Qy 15 ThGlnIlePhesys-----HisAsnAspTyrProGluThrIleThr 28
Db 205 TCTCACTGGCCGCTATTCATCTTACGCTTAACCTCACCTCGCTGCTCTAATACAAC 264
Qy 29 AspTyrValThrLeuGlnArg---GlySerAlaTyrGlyValLeuSerAsnPheser 47
```

RESULT 41
AL666950/c

LOCUS	AL6866950	826 bp	MRNA	linear	EST 11-JAN-2002
DEFINITION	AL6866950	directional larval cDNA	library	Clona intestinalis cDNA	
ACCESSION	clone 0132C09 5'	mRNA sequence.			
	AL6866950				

```

VERSION      AL666550.1  GI:18133857
KEYWORDS     EST.
SOURCE       .

```

ORGANISM
Clona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cloniidae. Clona

AUTHORS	Genoscope.
TITLE	Clona intestinalis directional larval cDNA library
JOURNAL	Unpublished (2002)
COMMENT	Contact: Genoscope

BP 191 91006 EVRY cedex - France
Email: segreff@genoscope.cns.fr, Web : www.genoscope.cns.fr
IMPORTANT: this sequence may contain errors. The Clona intestinalis library from which the clone was isolated may be contaminated with cDNAs from bacteria or other Eukarya.
Directional larval cDNA library originate from Dr. M.Branno, Stazione A.Dohrn, Naples, Italy, and was prepared in pBluescriptSK+.

```

source
1. .826

```

BASE COUNT	ORIGIN
204 a	247 c 160 g 213 t 2 others

Alignment Scores:	
Pred. No.:	
Score:	36.7
Percent Similarity:	81.00
Best Local Similarity:	40.318
Query Match:	27.918
8:	9.558
9	
Length:	826
Matches:	36
Conservative:	61
Mismatches:	16
Indels:	17
Gaps:	4

US-09-900-575-29_COPY_26_186 (1-161) x AL666950 (1-826)

38 AlaTyrGlyGlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyr 57

RESULT 42
AL667437/C

LOCUS	AL667437	868 bp	mRNA	linear	EST 11-JAN-2002
DEFINITION	AL667437	directional larval cDNA	library	clona	intestinalis cDNA
ACCESSION	clone 0192E10 5'	mRNA sequence.			
	AL667437				

```

VERSION      2.00/437.1      GI:18134344
KEYWORDS
SOURCE
Ciona intestinalis.

```

REFERENCE

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiaceae; Enterogona; Phlebobranchia; Clonidae; Clona.
1 (bases 1 to 868)

AUTHORS	Genoscope, Genoscope, 1 rue de l'Université, 75013 Paris, France
TITLE	Clona intestinalis directional larval cDNA library
JOURNAL	Unpublished (2002)
COMMENT	Contact: Genoscope

BR 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
IMPORTANT: this sequence may contain errors. The Clona Intestinales library from which the clone was isolated may be contaminated with cDNAs from bacteria or other Eukarya.
Directional larval cDNA library originate from Dr.M.Branno, Strazzone A.Dohn, Naples, Italy, and was prepared in pBluescriptSK+.

```

source      1. .868

```

```

/organism="Clostridium"
/db_xref="taxon:7719"
/clone_id="019ZE10"
/clone_lib="directional larval CDNA library"
/notes="Vector: pBluescript2SK+"
BASE COUNT      224 a      269 c      168 g      204 t      3 others
ORIGIN

```

Alignment Scores:	
Pred. No.:	39.7
Score:	81.00
Percent Similarity:	40.31%
Best Local Similarity:	27.91%
Query Match:	9.55%
8:	9
Length:	868
Matches:	36
Conservative:	61
Mismatches:	16
Indels:	17
Gaps:	4

US-09-900-575-29_COPY_26_186 (1-161) x AL667437 (1-868)

```

Oy 38 AlaTyrGlyValLeuSerAspNheSerGlyThrValLysTyrSerGlySerSerTyr 57
    ||| ||||| ||||| |||
Db 586 GGTGATGGCGGCGACATCTAATACGATGATGTAATTGGTCAAAATGACCGCTTG 527
Oy 58 PropheProThrThrSerGluThrProArgValValTyrAsnSer----- 72
    ||| ||||| ||||| |||
Db 526 CACTGGCAGATTGTGAAGTGTCTCCGGCAACGATATATCCAGTTGGCATTTGGTATGTA 467
Oy 73 ArgThrAspLysProThrProValAlaLeuThrLeuThrProValSerSerAlaGlyGly 92
    ||| ||||| ||||| |||
Db 466 AAAGCTGATCCAACTGCTGACTGGCGGAGCAAGAACGCCAAATGCTGTACAGGGGG 407
Oy 93 LeuVal-----lIeLysAlaGlySerLeuIleAlaValLeuLeuArg 107
    ||||| ||||| ||||| |||
Db 406 TTGGTGCAAGTAACTAATTAATCACTGCGGTGGTTAGTGTCACTCGAACCAAG-CGT 348
Oy 108 GluThrAsnAsnTyrAsnSerAspAspPheGlnPheValTyrPasnIleTyrAlaAsn 127
    ||| ||||| ||||| |||
Db 347 GCGGAAGGAGCACTGGGAGCAACTGGTGCATGTGACTGGAACTCTTGGCAGAAATCA 288
Oy 128 AspValValAlaProThrGlyGlyCysAspValSer-----AlaArgAspValThr 144
    ||| ||||| ||||| |||
Db 287 GACAGATTT-----GGTTGGTGTGAATGTCCAGTTCTCTGGCAGATGACTTGGG 237
Oy 145 ValThrLeuProAspTyrArgGlySer 153
    ||||| ||||| ||||| |||
Db 236 ATTCTCTGATGACGTAAGGCGCTCG 210

RESULT 43
Bi208686/c 675 bp mRNA linear EST 11-JUL-2001
LOCUS EST526726 cTOS Lycopersicon esculentum cDNA clone cTOS17H8 5' end,
DEFINITION mRNA sequence.
ACCESSION Bi208686
VERSION Bi208686.1 GI:14686410
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 675)
AUTHORS van der Hoeven, R., Sun, H., Bezzerides, J., Cho, J., Utterback, R.,
Rommig, C. and Tanksley, S.
Generation of ESTs from Tomato Suspension Cultures
Unpublished (2001)
CONTACT: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
Location/Qualifiers
1. 675
/organism="Lycopersicon esculentum"
/cultivar="TA496, B6203"
/db_xref="taxon:4081"
/clone_id="cTOS17H8"
/clone_lib="cTOS"
/tissue_type="suspension cultures"
/lab_host="SOLR"
/feature="Vector: pBluescript SK(-); site_1: EcoRI; site_2:
XhoI; Suspension cultures of L.esculentum E6203 were grown
in Murashige and Skoog based medium, supplemented with 15%
coconut milk (filter sterilized and added after
autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
Fresh medium was added every 7 days, and cultures were
grown at 25 C, with 12hrs of light and continuous
shaking."

BASE COUNT 222 a 97 c 177 g 179 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 30.8 Length: 675
Score: 80.50 Matches: 33
Percent Similarity: 35.62% Conservative: 19
Best Local Similarity: 22.60% Mismatches: 53
Query Match: 9.49% Indels: 41
DB: 13 Gaps: 5

```

US-09-900-575-29_COPY_26_186 (1-161) x Bi208686 (1-675)

```

Oy 15 ThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGln 34
    ||| ||||| ||||| |||
Db 633 ACTAATCAATATTTCCATCCCAATATCAAAATCACTACTCTTCTTCTTTGTTCAA 574
Oy 35 ArgGlySerAlaTyrGlyValLeuSerAspNheSerGlyThrValLysTyrSerGly 54
    ||| ||||| ||||| |||
Db 573 CAACGCTATTAAGGGGGAATCATTT----- 547
Oy 55 SerSerTyrPropheProThrThrSerGluThrProArgValValTyrAsnSerArgThr 74
    ||||| ||||| ||||| |||
Db 546 -----CCAACTCAAAATTTTCCGAACCTCTTCTGCTGTAGTTCTCTCTCT 502
Oy 75 AspLysProThrProValAlaLeuThrLeuThrProValSerSerAlaGlyGlyLeuVal 94
    ||| ||||| ||||| |||
Db 501 GATAACCCACAGATTTGTCCATGCTGATTT-----GGTAAAGAGCGCTTCTG 454
Oy 95 IleLysAlaGlySerLeuIleAlaValLeu-----lIeLeu 106
    ||||| ||||| ||||| |||
Db 453 TTC-----TCTTACTAGCTGTGCTTCAAGAAATCACTTGAAGTTCTCTCTC 403
Oy 107 ArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTyrPasnIleTyrAlaAsn 126
    ||| ||||| ||||| |||
Db 402 TCTGCTATTGTGTGAGAACCAACACTGATTTCTAGCGCTTGGACATCTTA----- 349
Oy 127 AsnAspValValAlaProThrGlyGlyCysAspValSerAlaArgAspValThrValThr 146
    ||| ||||| ||||| |||
Db 348 -----GGCTCTCTCTTTTCTTCTGGAAGATTCAACTACTACT 313
Oy 147 LeuProAspTyrArgGly 152
    ||| ||||| ||||| |||
Db 312 CCCAATCATTCACCTGCT 295

RESULT 44
Bi208424/c 731 bp mRNA linear EST 11-JUL-2001
LOCUS EST526464 cTOS Lycopersicon esculentum cDNA clone cTOS17E9 5' end,
DEFINITION mRNA sequence.
ACCESSION Bi208424
VERSION Bi208424.1 GI:14686148
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 731)
AUTHORS van der Hoeven, R., Sun, H., Bezzerides, J., Cho, J., Utterback, R.,
Rommig, C. and Tanksley, S.
Generation of ESTs from Tomato Suspension Cultures
Unpublished (2001)
CONTACT: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
Location/Qualifiers
1. 731
/organism="Lycopersicon esculentum"
/cultivar="TA496, B6203"
/db_xref="taxon:4081"
/clone_id="cTOS17E9"
/clone_lib="cTOS"
/tissue_type="suspension cultures"

```

```

/lab_host="SOLR"
/Note="Vector: pBluescript SK(-); Site: 1: EcoRI; Site: 2:
XhoI; Suspension cultures of L. esculentum E6203 were grown
in Murashige and Skoog based medium, supplemented with 15%
coconut milk (filter sterilized and added after
autoclaving), 2% saccharose, and 1mg/ml 2,4d (pH 5.8).
Fresh medium was added every 7 days, and cultures were
grown at 25 C, with 12hrs of light and continuous
shaking."
BASE COUNT      235 a      108 c      192 g      196 t
ORIGIN

Alignment Scores:
Pred. No.:      34.9      Length:      731
Score:          80.50      Matches:      33
Percent Similarity: 35.62%      Conservative: 19
Best Local Similarity: 22.60%      Mismatches: 53
Query Match:      9.49%      Indels:      41
DB:              13      Gaps:      5

US-09-900-575-29_COPY_26_186 (1-161) x BI208424 (1-731)
QY      15  ThGlnIlePheCysHIsAsnAspTYrProGluThrIleThrAspTYrValThrLeuGln 34
      |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      633 ACTAAGTACATTTGCCATCCAAATCATCCAAAGAACTACTACTCTTCTTCTTCTTCTTCA 574
QY      35  ArgGlySerAlaTYrGlyGlyValLeuSerAsnPheSerGlyThrValIstYrSerGly 54
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      573 CAAACGCTCTATAGGGGGGAAATCATTT----- 547
QY      55  SerSerTYrProPheProThrSerGluThrProArgValAlaTYrAsnSerArgThr 74
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      546 -----CCAACTCAAAATTTTCGAACTCTTCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 502
QY      75  AspLYsProTrpProValAlaLeuTYrLeuThrProValSerSerAlaGlyLeuVal 94
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      501 GATTAACCCCAAGATTTGTCATGCTGAT-----GGTAAGGAGGCTCTCTG 454
QY      95  IleLYsAlaGlySerLeuIleAlaValLeu-----IleLeu 106
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      453 TTC-----TCCCTTACTAGCTGTGCTCTCAAGAAATCACTGAAAGTTCTCTCTC 403
QY      107  ArgGlnThrAsnAsnTYrAsnSerAspPheGlnPheValITrPAsnIleTYrAlaAsn 126
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      402 TCTGCTAATGCTGTAGAGCAACACGATTTTCCAGGCTTTGCAACATCTTA----- 349
QY      127  AsnAspValValValProThrGlyGlyCysAspValSerAlaArgAspValThrValThr 146
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      348 -----GGCTCTCTCTTTTCTTCTGAAAGATTCACATCTCACT 313
QY      147  LeuProAspTYrArgGly 152
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      312 CCCAAATCATCACTGCT 295

RESULT 45
LOCUS      BG428847              779 bp      mRNA      linear      EST 14-MAR-2001
DEFINITION 602494632F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4608631 5',
            mRNA sequence.
ACCESSION  BG428847
VERSION    BG428847.1  GI:13335353
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.

```

```

cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LICM1350 row: P column: 08
High quality sequence stop: 673.
FEATURES
SOURCE      Location/Qualifiers
            1..779
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4608631"
            /clone_1lb="NIH_MGC_75"
            /lab_host="DH10B (T1 phage-resistant)"
            /note="Organ: Kidney; Vector: pDNR-LIB (Clontech); Site: 1:
            SfiI (ggcgccgcggcc); Site: 2: SfiI (ggccatcattggcc); 5' and
            3' adaptors were used in cloning as follows: 5' adaptor
            sequence: 5'-CACGGCCATATGCGC-3' and 3' adaptor sequence:
            5'-ATTCTAGAGCGCGAGCGCGCCGACATG-dt(30)BN-3' (where B = A,
            C, or G and N = A, C, G or T). Average insert size 1.65
            kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
            by PCR. This library was enriched for full-length clones
            and was constructed by Clontech Laboratories (Palo Alto,
            CA). Note: this is a NIH_MGC Library."
BASE COUNT      212 a      166 c      204 g      197 t
ORIGIN

Alignment Scores:
Pred. No.:      38.5      Length:      779
Score:          80.50      Matches:      31
Percent Similarity: 36.84%      Conservative: 18
Best Local Similarity: 23.31%      Mismatches: 31
Query Match:      9.49%      Indels:      53
DB:              12      Gaps:      6

US-09-900-575-29_COPY_26_186 (1-161) x BG428847 (1-779)
QY      13  LeuSerThGlnIlePheCysHIsAsnAspTYrProGluThrIleThrAspTYrValThr 32
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      131 CTGACTCCCTAATTTAACTTGAAGCCGCTGCTCTTCCAGGCGCTCACTTAC 190
QY      33  Leu-----GlnArgLYSerAlaTYrGlyGlyValLeuSerAsnPheSerGlyThrVal 50
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      191 CTCATGATGAAGAAAGGAGGAGGCTGAGACAGCTGTGTTCAAT----- 235
QY      51  LysTYrSerGlySerSerTYrProPheProThrThrSerGluThrProArgValAlaTYr 70
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      236 -----CCAAGCTCACCACTTCTTCACAGTAATCACTGACATA 274
QY      71  AsnSerArg-----ThrAspLYsProTrpProVal 80
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      275 AATGGAGAGCACTTAAGGCCCCCAAAACACTGTGAGTGAAGAAAGAACCTTACCCTCA 334
QY      81  AlaLeuTYrLeuThrProValSerSerAlaGlyGlyValIleLYsAlaGlySerLeu 100
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      334 ----- 334
QY      101  IleAlaValLeuIleLeuArgGlnThrAsnAsnTYrAsnSerAspPheGlnPhe--- 119
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      335 ---GCAGTGTGCTGTC-----TCCAGTGAATTAATCTGTGATGATCCCACTTTACC 382
QY      120 -----ValTrPAsnIleTYrAlaAsnAsnsp 128
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      383 GCTTGTGTAATCTGAATCTGAAGGCTATGCAATGAT 421

```

Search completed: November 28, 2002, 20:31:09
 Job time : 1986 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 28, 2002, 19:06:11 ; Search time 257 Seconds

(without alignments)
1410.785 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186

Perfect score: 848
Sequence: 1 PVTNVCNQLVVDISTQIFCH.....DVTVTLPDYGVSPIPLTVY 161

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DEV=x1h
-Q=/cgr2.1/USPTO.spool/US09900575/rnat.22112002.130708.4535/app.query.fasta.1.327
-DB=N_Geneseq.101002 -OPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR.SCORE=ptc -THR.MAX=100 -THR.MIN=0 -ALIGN=45
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09900575.cgrn.1.1.125.grnat.22112002.130708.4535 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N_Geneseq.101002.*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	837	24 AAD29358	Escherichia coli s
2	841	99.2	837	24 AAD29358	Escherichia coli s
3	841	99.2	900	16 AAD29358	Film protein gene
4	834	98.3	837	24 AAD29357	Escherichia coli s
5	834	98.3	837	24 AAD29357	Escherichia coli s
6	834	98.3	840	24 AAD29361	Escherichia coli s
7	834	98.3	840	24 AAD29361	Escherichia coli s
8	834	98.3	900	16 AAD29361	Film protein gene
9	834	98.3	903	21 AAD29361	Film protein gene
10	834	98.3	903	21 AAD29361	Film protein gene
11	834	98.3	903	21 AAD29361	Film protein gene
12	834	98.3	903	21 AAD29361	Film protein gene
13	834	98.3	4637	24 AAD29376	DNA encoding Esche
14	834	98.3	7416	24 AAD29374	Vector PF1MAGH ex
15	834	98.3	9299	24 AAD29374	Plasmid pCGA126-1
16	831	98.0	837	24 AAD29382	Vector PF1MAGH
17	830	97.9	837	24 AAD29382	Escherichia coli s
18	830	97.9	900	16 AAD29382	Escherichia coli s
19	826	97.4	837	24 AAD29382	Film protein gene
20	825	97.3	840	24 AAD29382	Escherichia coli s
21	825	97.3	900	16 AAD29382	Escherichia coli s
22	825	97.3	900	16 AAD29382	Film protein gene
23	825	97.3	900	16 AAD29382	Film protein gene
24	824	97.2	837	24 AAD29371	Escherichia coli s
25	824	97.2	837	24 AAD29371	Escherichia coli s
26	822	96.9	837	24 AAD29367	Escherichia coli s
27	822	96.9	837	24 AAD29367	Escherichia coli s
28	820	96.7	900	16 AAD29367	Escherichia coli s
29	820	96.7	837	24 AAD29367	Film protein gene
30	819	96.6	837	24 AAD29362	Escherichia coli s
31	819	96.6	837	24 AAD29362	Escherichia coli s
32	819	96.6	837	24 AAD29362	Escherichia coli s
33	819	96.6	840	24 AAD29365	Escherichia coli s
34	818	96.5	840	24 AAD29365	Escherichia coli s
35	818	96.5	837	24 AAD29365	Escherichia coli s
36	816	96.2	837	24 AAD29365	Escherichia coli s
37	816	96.2	837	24 AAD29365	Escherichia coli s
38	816	96.2	1227	24 AAD29365	Fusion protein prt
39	815	96.1	900	16 AAD29365	Film protein gene
40	815	96.1	900	16 AAD29365	Film protein gene
41	814	96.0	900	16 AAD29365	Film protein gene
42	814	96.0	900	16 AAD29365	Film protein gene
43	812	95.8	837	24 AAD29370	Escherichia coli s
44	812	95.8	837	24 AAD29370	Escherichia coli s
45	807	95.2	888	16 AAD29369	Film protein gene

ALIGNMENTS

RESULT 1
AAD29358
ID AAD29358 standard; DNA: 837 BP.

AC AAD29358:
XX 07-MAY-2002 (first entry)

DE Escherichia coli strain B240 Film DNA.

KW Film; Immune response; antibacterial; enterobacillus-related disease;
KW Therapy; vaccine; urinary tract infection; bladder; ds.

XX Escherichia coli B240.

OS
XX
FH Key Location/Qualifiers
FT CDS 1..837
FT /tag= a

```

FT      /product= "E. coli FimH protein"
FT      /transl_except= (pos:601..603,aa:His)
FT      /note= "CDS does not include start and stop codon"
FT      /partial
XX      W0200204496-A2.
XX      17-JAN-2002.
XX      06-JUL-2001; 2001WO-US21525.
XX      07-JUL-2000; 2000US-216750P.
XX      (MED1-) MEDIMMUNE INC.
XX      Langermann S, Revel A, Auguste C, Burtlein J;
XX      WPI: 2002-171702/22.
XX      P-PSDB: AAE18419.
XX      New immunogenic polypeptide, useful as vaccine for protecting against
XX      an enterobacillus-related disease in a patient at risk of contracting
XX      such disease, e.g. urinary tract infection or a bladder infection
XX      Example 1; Fig 1; 101pp: English.
XX      The invention relates to bacterial immunogenic agents for administration
XX      to humans and non-human animals to stimulate an immune response. The
XX      invention also relates to methods for vaccination of mammalian species
XX      with variants of E. coli FimH protein derived from different strains of
XX      E. coli. The vaccine composition or the antibody is useful for protecting
XX      against and creating an enterobacillus-related disease in a patient
XX      afflicted or at a risk of contracting the disease. In particular, the
XX      disease is a urinary tract or bladder infection. The disease is caused
XX      by a bacterium of the family Enterobacteriaceae, particularly E. coli.
XX      The present sequence is Escherichia coli strain B240 FimH DNA.
XX      Sequence 837 BP; 193 A; 200 C; 220 G; 224 T; 0 other:
SQ
Alignment Scores:
Pred. No.: 4,05e-91 Length: 837
Score: 848.00 Matches: 161
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0
US-09-900-575-29_COPY_26_186 (1-161) x AAD29358 (1-837)
QY      1 P-ovalValAsnValGlyGlnAsnLeuValAlaSplSerThGlnIlePheCysHis 20
QY      76 CCCGTCGTAATGTCGGAACCAACCGTCGTCGATCTTTGCACCCAAATCTTTGCCAT 135
DB      21 AsnAspTyPrcGluThrIleThrAspTyValThrLeuGlnArgGlySerAlaTyrcly 40
DB      136 AACGATATCCGGAAACCACTTACAGACTATGTCACACTCAGACGAGCGCTTATGAC 195
QY      41 GlyValLeuSerAsnPheserGlyThrValIlyTySerglySerSerTyrrProPhePro 60
DB      196 GCGCGTATCTAATTTTCCGGACCGCTAAATATAGGAGCAAGTATCCATTCTCT 255
QY      61 ThrThrSerGluThrProArgValAlaTyrrAsnSerArgThAspIysPrcPrrProVal 80
DB      256 ACCACACGGAACCGCCGCGCTGTTTATTAATTCAGAGAACGATTAAGCCGTCGCGGTC 315
QY      81 AlaLeuTyrrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
DB      316 GCGCTTATTTGACCGCTGTCGACAGTCGCGGCGGCTGTCGATTAACCTGCTCATTA 375
QY      101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrrAsnSerAspAspPheGlnPheVal 120
DB      376 ATTCGCGTCTTATTTTGGACACCAACACTATTAACAGCATGATTTCCAGTTTGG 435

```

```

QY      121 TrpAsnIleTyrrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB      436 TGGAAATATTTACGCCAATATGATGTGTGCTTACTGGCGGCTCGCATGTTTCGCT 495
QY      141 ArgAspValThrValThrLeuProAspTyrrArgGlySerValProIleProLeuThrVal 160
DB      496 CGTGATGTACACGTTACTCTCCGCACTACCGTTCAGTGCAGTTCCTCTTACCGTT 555
QY      161 Tyr 161
DB      556 TAT 558
RESULT 2
AAD29368
ID      AAD29368 standard; DNA: 837 BP.
XX      AAD29368:
XX      07-MAY-2002 (first entry)
XX      Escherichia coli strain EC80 FimH DNA.
XX      FimH; immune response; antibacterial; enterobacillus-related disease;
XX      therapy; vaccine; urinary tract infection; bladder; ds.
XX      Escherichia coli EC80.
XX      Key Location/Qualifiers
XX      CDS 1..837
XX      /*tag= a
XX      /product= "E. coli FimH protein"
XX      /transl_except= (pos:79..81,aa:Ala)
XX      /transl_except= (pos:493..495,aa:Val)
XX      /transl_except= (pos:511..513,aa:Ile)
XX      /transl_except= (pos:805..807,aa:Lys)
XX      /note= "CDS does not include start and stop codon"
XX      /partial
XX      W0200204496-A2.
XX      17-JAN-2002.
XX      06-JUL-2001; 2001WO-US21525.
XX      07-JUL-2000; 2000US-216750P.
XX      (MED1-) MEDIMMUNE INC.
XX      Langermann S, Revel A, Auguste C, Burtlein J;
XX      WPI: 2002-171702/22.
XX      P-PSDB: AAE18429.
XX      New immunogenic polypeptide, useful as vaccine for protecting against
XX      an enterobacillus-related disease in a patient at risk of contracting
XX      such disease, e.g. urinary tract infection or a bladder infection
XX      Example 1; Fig 1; 101pp: English.
XX      The invention relates to bacterial immunogenic agents for administration
XX      to humans and non-human animals to stimulate an immune response. The
XX      invention also relates to methods for vaccination of mammalian species
XX      with variants of E. coli FimH protein derived from different strains of
XX      E. coli. The vaccine composition or the antibody is useful for protecting
XX      against and creating an enterobacillus-related disease in a patient
XX      afflicted or at a risk of contracting the disease. In particular, the
XX      disease is a urinary tract or bladder infection. The disease is caused
XX      by a bacterium of the family Enterobacteriaceae, particularly E. coli.
XX      The present sequence is Escherichia coli strain EC80 FimH DNA.
XX      Sequence 837 BP; 191 A; 200 C; 223 G; 223 T; 0 other:
SQ
Alignment Scores:

```

Pred. No.: 2.75e-90 Length: 837
 Score: 841.00 Matches: 159
 Percent Similarity: 99.38% Conservative: 1
 Best Local Similarity: 98.76% Mismatches: 1
 Query Match: 99.17% Indels: 0
 DB: 24 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29368 (1-837)

QY 1 ProValValAsnValGlyGlnAsnLeuValAlaPleuSerThrGlnIlePheCysHis 20
 DB 76 CCGCGTGAATGTGGGCGCAAAACCTGCTGCTGATCTTTGACGCAATCTTTGGCAT 135
 QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
 DB 136 AACGATTAATCCGGAACCATTAACAGACTATGTACACCTGCAACGAGCGCTGATGGC 195
 QY 41 GlyValLeuSerAsnPheSerGlyThrValIleTyrSerGlySerSerTyrProPhePro 60
 DB 196 GCGGTATATCTATTTTTCGGGACCGCTAATAATATGTGGCACTAGCTATCCATTTGCT 255
 QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
 DB 256 ACCACGACGCAACGCGCGCTGTTTAAATTCGAGAACGGATAGCCGTGCGCGT 315
 QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
 DB 316 GCGCTTATTTTACGCGCTGTGAGAGTGGCGGGGGGCGATTTAAAGCTGCTCATTTA 375
 QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
 DB 376 ATTGCGCTGCTTATTTTGGCAGACCAACACTATTAACAGCATGATTTCCAGTTTGG 435
 QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
 DB 436 TGGAAATTTTACGCCAATAATATGTGTGCTCTACTGCGCGTGCATTTGCTGCT 495
 QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
 DB 496 CGTGATGTCTACCGTTACTCTGCGGACTACCGTGTTCAGTTCAGTTCCTCTTACCGTT 555
 QY 161 Tyr 161
 DB 556 TAT 558

RESULT 3
 AAQ93062
 ID AAQ93062 standard; DNA; 900 BP.
 XX
 AC AAQ93062;
 DT 15-MAR-1996 (first entry)
 XX
 DE FimH protein gene from E. coli #1.
 XX
 KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
 KW FimH; FimH; receptor binding site; ss.
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT sig_peptide 1..63
 FT /*tag= a 64..900
 FT mat_peptide /*tag= b
 XX
 PD 03-AUG-1995.
 XX
 PF 27-JAN-1995; 95W0-DK00042.
 XX
 PR 27-JAN-1994; 94US-0187166.

XX
 PA (GXBI-) GX BIOSYSTEMS AS.
 XX
 PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
 XX
 DR WPI; 1995-275442/36.
 XX
 PT Receptor specific bacterial adhesins - useful for targeting active
 PT compounds and microbial cells to locations of receptors
 PS Example 5; Page 114; 152pp; English.

The sequences given in AAQ93062-75 encode FimH proteins from various E. coli clinical isolates. FimH is located at the tip of the type 1 fimbriae and also intercalated at intervals in the fimbrial organelle. Most forms of the FimH adhesin target to, and bind to, oligosaccharide structures containing terminally located alpha-D-mannoside residues. CC FimH contains 4 cysteine residues assumed to direct folding of the CC molecule into distinct functional domains. For comparison FimA and CC the minor components FimF and FimG only have 2 cysteine residues. CC The localisation of the cysteine residues in FimH points to a tandem CC arrangement of two ancestral genes. Similar amino acids can be CC found in similar positions in the two halves of the FimH protein. The CC "midway" point is located roughly around residue 150 in the mature CC binding site, whereas the C-terminal sector became the receptor CC molecule required for integration into the fimbrial organelle. These CC sequences may be used in the production of variant FimH adhesins which CC may be useful for targeting active compounds and microbial cells to CC locations comprising selected receptors to which the adhesins bind. CC

SQ Sequence 900 BP; 204 A; 212 C; 241 G; 243 T; 0 other;

Alignment Scores:

Pred. No.: 3.05e-90 Length: 900
 Score: 841.00 Matches: 159
 Percent Similarity: 99.38% Conservative: 1
 Best Local Similarity: 98.76% Mismatches: 1
 Query Match: 99.17% Indels: 0
 DB: 16 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AAQ93062 (1-900)

QY 1 ProValValAsnValGlyGlnAsnLeuValAlaPleuSerThrGlnIlePheCysHis 20
 DB 139 CCGCGTGAATGTGGGCGCAAAACCTGCTGCTGATCTTTGACGCAATCTTTGGCAT 198
 QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
 DB 199 AACGATTAATCCGGAACCATTAACAGACTATGTACACCTGCAACGAGCGCTGATGGC 258
 QY 41 GlyValLeuSerAsnPheSerGlyThrValIleTyrSerGlySerSerTyrProPhePro 60
 DB 259 GCGGTATATCTATTTTTCGGGACCGTAAATAATGTGCGCATCTATTCATTTGCT 318
 QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
 DB 319 ACCACGACGCAACGCGCGCTGTTTAAATTCGAGAACGATTAACCGCTGCGGTG 378
 QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
 DB 379 GCGCTTATTTTACGCGCTGTGAGAGTGGCGGGGGGCGATTTAAAGCTGCTCATTTA 438
 QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
 DB 439 ATTGCGCTGCTTATTTTTCGACAGACCAACACTATTAACGATGATTTCCAGTTTGG 498
 QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
 DB 499 TGGAAATTTTACGCCAATAATATGTGTGCTGCTACTGCGGCGTGCATGTTTCTGCT 558
 QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160

```

|||||
Db 559 CGTATGTACACGTTACTGCGGACTACGTTGTTCACTGCAATTCCTTACCGTT 618
QY 161 Tyr 161
Db 619 TAT 621

RESULT 4
AAD29357
ID AAD29357 standard; DNA: 837 BP.
XX
AC AAD29357;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain B238 FimH DNA.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli B238.
XX
FH key
FH CDS location/Qualifiers
FT 1..837
FT /*tag= a
FT /product= "E. coli FimH protein"
FT /transl_except= (pos:70..72, aa:Leu)
FT /transl_except= (pos:76..78, aa:Ile)
FT /transl_except= (pos:526..528, aa:Arg)
FT /transl_except= (pos:601..603, aa:His)
FT /transl_except= (pos:817..819, aa:Gly)
FT /transl_except= (pos:820..822, aa:Ala)
FT /note= "CDS does not include start and stop codon"
FT /partial
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burteln J;
XX
DR WPI: 2002-171702/22.
DR P-PSDB: AAE18418.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
XX
PS Example 1; Fig 1; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain B238 FimH DNA.
XX
SQ Sequence 837 BP; 192 A; 202 C; 223 G; 220 T; 0 other;

```

```

Best Local Similarity: 98.14% Mismatches: 2
Query Match: 98.35% Indels: 0
DB: 24 Gaps: 0
US-09-900-575-29_copy_26_186 (1-161) x AAD29357 (1-837)
QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
Db 76 CCCGCCGTGAATGTGGGAGAAACCTGCTGATCTTTCAGCAATCTTTGCCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerLeuTyrGly 40
Db 136 AACGATTACCCGGAACCATTAAGATTAATGTCACACTGCAACGAGGCTGGCTTATGCG 195
QY 41 GlyValLeuSerAsnProSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
Db 196 GGCCTGTACTATATTTTTCGGGACCGTAAATATATAGTGGAGATGATCATTTTCG 255
QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProThrProVal 80
Db 256 ACCACCAAGTGAACGGCGCGGCTGTTTAAATTCAGAAAGGATTAAGCCGTCGGGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
Db 316 CGCCTTATTTGACCGCTGTGACAGTGGCGGGGTGTGATTAACCTGGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
Db 376 ATTCGCTGCTTATTTTTCGACAGACCAACATATTAACGATATTTCCAGTTGTG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValIleProThrGlyGlyAspValSerAla 140
Db 436 TGGAAATTTTACGCCAATATGATGTGTGTGTCGCCACATGCGCGGCGCATGTTTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
Db 496 CGTATGTACACGTTACTCTGCGGACTACCTGTGTTCAATGCGAATCTCTTACCGTT 555
QY 161 Tyr 161
Db 556 TAT 558

RESULT 5
AAD29372
ID AAD29372 standard; DNA: 837 BP.
XX
AC AAD29372;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain J96 FimH DNA.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli J96.
XX
FH key
FH CDS location/Qualifiers
FT 1..837
FT /*tag= a
FT /product= "E. coli FimH protein"
FT /note= "CDS does not include start and stop codon"
FT /partial
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMUNE INC.

```

Alignment Scores: 1.87e-89 Length: 837
 Score: 834.00 Matches: 158
 Percent Similarity: 98.76% Conservative: 1

XX Langermann S, Revel A, Auguste C, Burteln J;
 XX WPI: 2002-171702/22.
 DR P-PSDB: AAE18433.
 XX
 PT New immunogenic polypeptide, useful as vaccine for protecting against
 PT an enterobacillus-related disease in a patient at risk of contracting
 XX such disease, e.g. urinary tract infection or a bladder infection
 XX
 PS Example 1; Fig 1; 101pp; English.
 XX
 CC The invention relates to bacterial immunogenic agents for administration
 CC to humans and non-human animals to stimulate an immune response. The
 CC invention also relates to methods for vaccination of mammalian species
 CC with variants of E. coli FimH protein derived from different strains of
 CC E. coli. The vaccine composition or the antibody is useful for protecting
 CC against and treating an enterobacillus-related disease in a patient
 CC afflicted or at a risk of contracting the disease. In particular, the
 CC disease is a urinary tract or bladder infection. The disease is caused
 CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
 CC The present sequence is Escherichia coli strain J96 FimH DNA.
 XX
 SQ Sequence 837 BP; 191 A; 201 C; 222 G; 223 T; 0 other:
 Alignment Scores:
 Pred. No.: 1.87e-89 Length: 837
 Score: 834.00 Matches: 158
 Percent Similarity: 98.76% Conservative: 1
 Best Local Similarity: 98.14% Mismatches: 2
 Query Match: 98.35% Indels: 0
 DB: 24 Gaps: 0
 US-09-900-575-29_COPY_26_186 (1-161) x AAD29361 (1-837)
 QY 1 ProValValAsnValGlyGlnAsnLeuValAlaAspLeuSerThrGlnIlePheCysHis 20
 DB 76 CCCGCGGAAATGGGGCAAAACCTGCTGATCTTTGACGCAAAATCTTTGGCAT 135
 QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
 DB 136 AACGATTAATCCGAAACCATTAACACTGACACTGCACACGAGCGCTTATGGC 195
 QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
 DB 196 GCGCTGATATCTTTTCCGGACCGTAAATATATGTCGACATCTCAATTCAT 255
 QY 61 ThrThrSerGluThrProArgValAlaTyrAsnSerArgThrAspLysProTyrProVal 80
 DB 256 ACCACGACGAAACGCGCGGTGTTATATCGAAGAACGATAGCCGTGCGCGGTG 315
 QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
 DB 316 GCGCTTATTTGACGCTGTGACAGTGCAGCGCGGTGCGCATTAACCTGCGCTCAT 375
 QY 101 IleAlaValLeuIleLeuArgGlnThrAsnTyrAsnTyrAsnSerAspPheGlnPheVal 120
 DB 376 ATGGCGCTGATTTTGGCAGACACCAACACTATTAACGCAATGATTTCCAGTTTGG 435
 QY 121 TrpAsnIleTyrAlaAsnAspValAlaValProThrGlyGlyCysAspValSerAla 140
 DB 436 TGGAAATTTTACCCCAATATATGATGTCGTCCTACTCGCGGTGCGCATTTCTGCT 495
 QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
 DB 496 CGTGATGTCCACCTTACTCTGCGGACTACCTGCTTCACTGCAATTCCTTTACCGTT 555
 QY 161 Tyr 161
 DB 556 TAT 558
 RESULT 6
 AAD29361

ID AAD29361 standard; DNA; 840 BP.
 AC AAD29361;
 XX
 DT 07-MAY-2002 (first entry)
 XX
 DE Escherichia coli strain EC42 FimH DNA.
 XX
 KW FimH; immune response; antibacterial; enterobacillus-related disease;
 KW therapy; vaccine; urinary tract infection; bladder; ds.
 XX
 OS Escherichia coli EC42.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..840
 FT /*tag= a
 FT /product= "E. coli FimH protein"
 FT /note= "CDS does not include start codon"
 XX
 PN MO200204496-A2.
 XX
 PD 17-JAN-2002.
 XX
 PF 06-JUL-2001; 2001MO-US21525.
 XX
 PR 07-JUL-2000; 2000US-216750P.
 XX
 PA (MEDI-) MEDIMUNE INC.
 XX
 PI Langermann S, Revel A, Auguste C, Burteln J;
 XX P-PSDB: AAE18422.
 DR WPI: 2002-171702/22.
 XX
 PT New immunogenic polypeptide, useful as vaccine for protecting against
 PT an enterobacillus-related disease in a patient at risk of contracting
 PT such disease, e.g. urinary tract infection or a bladder infection
 XX
 PS Example 1; Fig 1; 101pp; English.
 XX
 CC The invention relates to bacterial immunogenic agents for administration
 CC to humans and non-human animals to stimulate an immune response. The
 CC invention also relates to methods for vaccination of mammalian species
 CC with variants of E. coli FimH protein derived from different strains of
 CC E. coli. The vaccine composition or the antibody is useful for protecting
 CC against and treating an enterobacillus-related disease in a patient
 CC afflicted or at a risk of contracting the disease. In particular, the
 CC disease is a urinary tract or bladder infection. The disease is caused
 CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
 CC The present sequence is Escherichia coli strain EC42 FimH DNA.
 XX
 SQ Sequence 840 BP; 193 A; 202 C; 224 G; 221 T; 0 other:
 Alignment Scores:
 Pred. No.: 1.88e-89 Length: 840
 Score: 834.00 Matches: 158
 Percent Similarity: 98.76% Conservative: 1
 Best Local Similarity: 98.14% Mismatches: 2
 Query Match: 98.35% Indels: 0
 DB: 24 Gaps: 0
 US-09-900-575-29_COPY_26_186 (1-161) x AAD29361 (1-840)
 QY 1 ProValValAsnValGlyGlnAsnLeuValAlaAspLeuSerThrGlnIlePheCysHis 20
 DB 76 CCCGCGGAAATGGGGCAAAACCTGCTGATCTTTGACGCAAAATCTTTGGCAT 135
 QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
 DB 136 AACGATTAATCCGAAACCATTAACACTGACACTGCACACGAGCGCTTATGGC 195
 QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60

```

|||||
Db 196 GGCCTGTTATCTATTTTCGGACCGGTAAATATAGTGCACCTACCTATCCATTCCG 255
QY 61 ThrSerGluThrProAlaValValIYrAsnSerArgThrAspLysProTTPProVal 80
Db 256 ACCACCACTGAAGACCGCGGGTGTATATATTCGAGAACGGATAGCCGTGCGCGTG 315
QY 81 AlaLeuTYrLeuThrProValSerSerAlaGlyLeuValIleuValIleuValSerLeu 100
Db 316 GCGCTTTATTGACCGCTGTAGCAGTCGCGGGGGTGGTGGATTAAGCTGGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTYrAsnSerSerAspAspPheGlnPheVal 120
Db 376 ATTGCCCGTCTATTTCGACAGACCAACAATATAACACCGATTCAGTTTCCAGTTTGTG 435
QY 121 TrpAsnIleTYrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
Db 436 TGGATATATTACGCCAAATAATGATGTGGTGGCCACTGGCGGCTGGCATGTTTCTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTYrArgGlySerValProIleProLeuThrVal 160
Db 496 CGTGATGTCACCGTTACTCTGCGGACATACCTGGTTCAGTGGCGATTCCTTACCGTT 555
QY 161 TYr 161
Db 556 TAT 558

RESULT 7
AAD29363
ID AAD29363 standard; DNA; 840 BP.
XX
AC AAD29363;
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC56 FimH DNA.
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli EC56.
XX
FH Key Location/Qualifiers
FT CDS 1..840
FT /tag= "a"
FT /product= "E. coli FimH protein"
FT /note= "CDS does not include start codon"
FT /partial

MO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001MO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MED-) MEDIMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burslein J;
XX
DR WPI: 2002-171702/22.
XX
DR P-PSDB; AAE18424.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
XX
PS Example 1; Fig 1; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species

```

```

CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC56 FimH DNA.
XX
SQ Sequence 840 BP; 193 A; 201 C; 222 G; 224 T; 0 other;
XX
Alignment Scores:
Pred. No.: 1,88e-89 Length: 840
Score: 834.00 Matches: 158
Percent Similarity: 98.76% Conservative: 1
Best Local Similarity: 98.14% Mismatches: 2
Query Match: 98.35% Indels: 0
DB: 24 Gaps: 0

US-09-900-575-29_copy_26_186 (1-161) x AAD29363 (1-840)
QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
Db 76 CCCGCTGATGATGGGGCAAACTGCTGTGATCTTCGACGCAATCTTTCCAT 135
QY 21 AsnAspTYrProGluThrIleThrAspTYrValThrLeuGlnArgGlySerAlaTYrGly 40
Db 136 AACGATTATCCGGAACCATTAAGACTATGTCACACTGCACAGAGCTCGCTTATGCG 195
QY 41 GlyValLeuSerAsnPheSerGlyThrValLysTYrSerGlySerTYrProPhePro 60
Db 196 GCGCTGTTATCTAAATTTTCGGGACCGTAATAATATGAGCACTATCTTCCATTTCCT 255
QY 61 ThrThrSerGluThrProAlaValValTYrAsnSerArgThrAspLysProTTPProVal 80
Db 256 ACCACCAAGCAACGCCGCGGCTGTTATAATTCGAGAACGATAGCCGTGCGCGTG 315
QY 81 AlaLeuTYrLeuThrProValSerSerAlaGlyLeuValIleuValIleuValSerLeu 100
Db 316 GCGCTTTATTGACCGCTGTAGCAGTCGCGGCGGGTGGGATTAAGCTGGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTYrAsnSerSerAspAspPheGlnPheVal 120
Db 376 ATTGCCCGTCTATTTCGACAGACCAACAATATAACACCGATTCAGTTTCCAGTTTGTG 435
QY 121 TrpAsnIleTYrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
Db 436 TGGATATATTACGCCAAATAATGATGTGGTGGCTACTGCGGCTGGCATGTTTCTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTYrArgGlySerValProIleProLeuThrVal 160
Db 496 CGTGATGTCACCGTTACTCTGCGGACATACCTGGTTCAGTGCATATTCCTTACCGTT 555
QY 161 TYr 161
Db 556 TAT 558

RESULT 8
AAQ93070
ID AAQ93070 standard; DNA; 900 BP.
XX
AC AAQ93070;
DT 15-MAR-1996 (first entry)
XX
DE FimH protein gene from E. coli #2.
XX
KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW FimH; FimF; FimG; receptor binding site; ss.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT sig_peptide 1..63

```

The sequences given in AA030362-75 encode FlmH proteins from various *E. coli* clinical isolates. FlmH is located at the tip of the type I fimbriae and also intercalated at intervals in the fimbrial organelle. Most forms of the FlmH adhered in target to, and bind to, oligosaccharide structures containing terminally located alpha-D-mannoside residues. FlmH contains 4 cysteine residues assumed to direct folding of the molecule into distinct functional domains. For comparison FlmA and the minor components FlmF and FlmG only have 2 cysteine residues. The localisation of the cysteine residues in FlmH points to a tandem arrangement of two ancestral genes. Similar amino acids can be found in similar positions in the two halves of the FlmH protein. The "midway" point is located roughly around residue 150 in the mature protein. The two halves or domains of FlmH have evolved differently with the N-terminal section becoming the domain harbouring the receptor binding site, whereas the C-terminal sector became the domain of the molecule required for integration into the fimbrial organelle. These sequences may be used in the production of variant FlmH adhesins which may be useful for targeting active compounds and microbial cells to locations comprising selected receptors to which the adhesins bind.

QY	1	ProValValaenValaGlyGlnAsnLeuValValaAspLeuSerTrgInIlePheCysHis	20
Dd	139	CCCGCGCGATGTGGGGCAAAACCTGGTCTGGATCTTTGACGCAAAATCTTTGCCAT	199
QY	21	AsnAspTrpProGlnThrIleThrAspTrpAlaThrLeuGlnAlaGlySerAlaTrpGly	40
Dd	199	AACGATTATCGGCAACACATTACAGACTATGTCACTGCAACAGAGCTCGGCTTAAGGC	258
QY	41	GlyValIleuSerAsnPheSerGlyThrValIstTrpSerGlySerTrpProPhePro	60
Dd	259	GGCGGTATTATATTTTTTCGGGACCGGTAAATAATGTGGCAATACCTATTCATTTCT	318
QY	61	ThrThrSerGlnThrProArgValValTrpAsnSerArgrThrAspLysProTrrProVal	80
Dd	319	ACCAACCAAGCAAAACGCGCGGTGTATTATATTCGAAACGGATAGCCGTGGCCGTC	378
QY	81	AlaLeuTrpLeuTrpProValaSerSerAlaGlyClyLeuValIleLysAlaGlySerLeu	100

The sequences given in AA0393062-75 encode FlmM, proteins from various *E. coli* clinical isolates. FlmM is located at the tip of the type 1 fimbriae and also intercalated at intervals in the fimbrial organelle. Most forms of the FlmM adhesin target to, and bind to, oligosaccharide structures containing terminally located alpha-D-mannoside residues. FlmM contains 4 cysteine residues assumed to direct folding of the molecule into distinct functional domains. For comparison FlmA and the minor components FlmF and FlmG only have 2 cysteine residues. The localisation of the cysteine residues in FlmM points to a tandem arrangement of two ancestral genes. Similar amino acids can be found in similar positions in the two halves of the FlmM protein. The "midway" point is located roughly around residue 150 in the mature protein. The two halves or domains of FlmM have evolved differently with the N-terminal section becoming the domain harbouring the receptor

|||||
DB 559 CGGTGATGCACCGTACTCTGCGGACATACCGTTCAGTGCATTCCTTACCGTT 618
OY 161 Tyr 161
|||
DB 619 TAT 621
RESULT 11
ABK1187
ID ABK1187 standard; DNA; 903 BP.
XX
AC ABK1187;
XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding Escherichia coli type I pilin protein, FimH.
XX
KW Immune response; primase; immunoglobulin; urogenital tract infection;
KW FimC-FimH complex; FimH; IgG; human; urinary tract infection; URT;
KW bladder infection; kidney infection; Enterobacteriaceae; bacteraemia;
KW pregnant woman; diabetic; immunocompromised; HIV; cancer;
KW human immunodeficiency virus infection; end stage renal disease;
KW type I pilin protein; adhesin; FimH; ds.
XX
OS Escherichia coli isolate J96.
XX
XX
FH Key
FT 1..903 Location/Qualifiers
FT /*tag= a
FT /product= "FimH"
XX
PN WC200215928-A1.
XX
PD 28-FEB-2002.
XX
PF 28-NOV-2000; 2000WC-US32398.
XX
PR 18-AUG-2000; 2000US-226146P.
XX
PA (MEDT-) MEDIMKUNE INC.
XX
PI Langermann S, Ballou WR;
XX
DR WPI: 2002-280859/32.
XX
DR P-PSDB: AAU77488.
XX
PT Stimulating immune response in a primate for preventing, treating
PT bacterial induced diseases such as diseases of urinary tract, by
PT administering bacterial adhesive proteins, preferably FimC-FimH
PT polypeptide complex.
XX
PS Disclosure: Page 88-89; 92pp; English.
XX
XX
CC The present invention relates to a method of inducing an immune response
CC in a primate. The response involves immunoglobulin (Ig) molecules that
CC bind a bacterial adhesin protein, preferably an attachment domain of
CC a type I pilin polypeptide (e.g. FimH) associated with a bacterium
CC causing urogenital tract infections (e.g. Escherichia coli). The method
CC comprises administering a purified FimH polypeptide, a FimC-FimH (FimCH)
CC complex, or immunogenic fragments of these. The method is useful for
CC inducing IgG molecules in a primate, especially human, to reduce or
CC prevent the incidence of urogenital tract infections, particularly
CC urinary tract infection (UTI), bladder infection, or kidney infection,
CC caused by a bacterium of the family Enterobacteriaceae, preferably
CC E. coli. The method can be used in a human subject that has suffered
CC more than two urogenital infections within one year, has asymptomatic
CC bacteraemia, is a pregnant woman or a diabetic, is immunocompromised, has
CC remission from cancer, or is at risk for end stage renal disease. The
CC method is useful for vaccinating a primate against urogenital tract
CC infections, for treating or ameliorating the symptoms of urogenital
CC tract infections, and also for slowing or preventing progression of
CC a urinary tract infection into end stage renal disease. The present

CC sequence encodes E. coli FimH protein.
XX
SQ Sequence 903 BP; 206 A; 213 C; 240 G; 244 T; 0 other:
Alignment Scores:
Score: 2.08e-89 Length: 903
Matches: 158
Percent Similarity: 98.76% Conservative: 1
Best Local Similarity: 98.14% Mismatches: 2
Query Match: 98.35% Indels: 0
DB: 24 Gaps: 0
US-09-900-575-29_COPY_26_186 (1-161) x ABK1187 (1-903)
OY 1 ProValValAsnValGlyGlnAsnLeuValAspLeuSerThrGlnIlePheCysHis 20
DB 139 CCCGTCGTGATGATGGGGCCAAACCTGTGTCGATCTTCCACGCAATCTTTGCCAT 198
OY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 199 AACGATTATCCGGAAACCATTAACAGACTATGTCACACTGCACACGAGCTCGCTTATGCG 258
OY 41 GlyValLeuSerAsnPheserGlyThrValIlyTyrSerGlySerSerTyrProPhePro 60
DB 259 GCGCTTATCTAATTTTCCGGACCGCTAAATATAGTGCAGTATCATTCATTCTCT 318
OY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProThrProVal 80
DB 319 ACCACACGCGAAACGCGCGCGCTGTTTATTAATTCAGACAGCATTAACCGTGGCGTG 378
OY 81 AlaIleuTyrLeuThrProValSerSerAlaGlyLeuValIleValAlaGlySerLeu 100
DB 379 GCGCTTATTTAGCCCTGTGACAGTGTGCGGCGCGCTGCGATTAAGCTGGCTCATTA 438
OY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
DB 439 ATTGCCGTGCTTATTTTGGACAGCACCAACACTATTAACAGCATTTCCACTTTGTG 498
OY 121 TTPAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 499 TGGAAATATTACGCCCAATATGATGTGCTGCTACTGCGGCTGCGATGTTTCTGCT 558
OY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 559 CGTGTATGTCACCGTACTCTCCGACATACCTGTGTCAATGTCATTCCTTACCGTT 618
OY 161 Tyr 161
DB 619 TAT 621
RESULT 12
AAD29376
ID AAD29376 standard; DNA; 903 BP.
XX
XX
AC AAD29376;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain J96 FimH DNA plus signal sequence.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli J96.
XX
PN WC200204496-A2.
XX
PD 17-JAN-2002.
XX
XX
PF 06-JUL-2001; 2001WC-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX

(MEDI-) MEDIMUNE INC.
 Langermann S, Revel A, Auguste C, Burlain J;
 WPI; 2002-11/1702/22.
 New immunogenic polypeptide, useful as vaccine for protecting against
 an enterobacillus-related disease in a patient at risk of contracting
 such disease, e.g. urinary tract infection or a bladder infection
 Example 2; Page 96; 101pp; English.
 The invention relates to bacterial immunogenic agents for administration
 to humans and non-human animals to stimulate an immune response. The
 invention also relates to methods for vaccination of mammalian species
 with variants of E. coli fimb protein derived from different strains of
 E. coli. The vaccine composition or the antibody is useful for protecting
 against and treating an enterobacillus-related disease in a patient
 afflicted or at a risk of contracting the disease. In particular, the
 disease is a urinary tract or bladder infection. The disease is caused
 by a bacterium of the family Enterobacteriaceae, particularly E. coli.
 The present sequence is Escherichia coli strain J96 fimb DNA plus signal
 sequence, used in the exemplification of the invention.
 Sequence 903 BP; 206 A; 213 C; 240 G; 244 T; 0 other:
 Alignment Scores:
 Pred. No.: 2 08e-89 Length: 903
 Score: 834.00 Matches: 158
 Percent Similarity: 98.76% Conservative: 1
 Best Local Similarity: 98.14% Mismatches: 2
 Query Match: 98.35% Indels: 0
 DB: 24 Gaps: 0
 US-09-900-575-29_COPY_26_186 (1-161) x AMD29376 (1-903)
 QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
 DB 139 CCGGTGTAATGTGGGGCAAAACCGTGGATCTTTCAGCAATCTTTGGCCAT 198
 QY 21 AsnAspTyrProGlnThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
 DB 199 AACGATTATCCGAAACCATTCACAGCTATGTCACACTGCACAGCGCGCTATATGSC 258
 QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
 DB 259 GGGGTATATCTAATTTTCCGGACCGTAAATATAGTGCGAGTATCCATTCTCT 318
 QY 61 ThrThrserGlnThrProArgValValTyrAsnSerArgThrAspLysProTrrProVal 80
 DB 319 ACCACCGCGAAACGGCGCGCTGTTTATTAATTCAGAAAGGATTAAGCGCGCGGTG 378
 QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
 DB 379 GCGGTTTATTTGACCGCTGTGACAGTGGCGGCGGCGGTAAAGCTGCTCATTA 438
 QY 101 IleAlaValLeuIleLeuArgGlnThrAsnTyrAsnSerAspPheGlnPheVal 120
 DB 439 ATTCCTGCTTATTTTGGACAGCAACCAACAATATTAACGCGATGATTTCCAGTTTGTG 498
 QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
 DB 499 TGGAAATATTTAGCCCAATATGATGTGTGCTACTGCGGCTCGATGTTTTCGCT 558
 QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
 DB 559 CGTATGTCACCGTTACTGTGCGGACTACCGTTCAGTCCCAATTCCTTACCGGTT 618
 QY 161 Tyr 161
 DB 619 TAT 621
 RESULT 13

ABK23921
 ID ABK23921 standard; DNA; 4637 BP.
 XX AC ABK23921;
 XX XX
 DT 09-APR-2002 (first entry)
 XX XX
 DE Vector pFIMAFGH expressing pilin genes fimF, G and H.
 XX XX
 KW Vaccine; molecular scaffold; pilus; pilin; HBcAg; antigen;
 KW hepatitis B virus capsid protein; JON; FOS; HIV gp140;
 KW measles virus N protein; bee venom phospholipase; Th type 2 T-helper;
 KW Th2; Sindbis virus E2 protein; amyloid beta; influenza M2 antigen;
 KW human immunodeficiency virus infection; viral hepatitis; measles;
 KW chicken pox; pneumonia; tuberculosis; syphilis; malaria; allergy;
 KW cancer; chronic disease; arthritis; colitis; diabetes;
 KW multiple sclerosis; ds.
 XX XX
 OS Escherichia coli.
 XX XX
 PN WO200185208-A2.
 XX XX
 PD 15-NOV-2001.
 XX XX
 PF 02-MAY-2001; 2001WO-IB00741.
 XX XX
 PR 05-MAY-2000; 2000US-202341P.
 XX XX
 PA (CYTO-) CYTOS BIOTECHNOLOGY AG.
 PA (SEBBEL) SEBBEL P.
 PA (DUNA) DUNANT N.
 PA (BACH) BACHMANN M.
 PA (TISSOT) TISSOT A.
 PA (LECHNER) LECHNER F.
 XX XX
 PI Sebbel P, Dunant N, Bachmann M, Tissot A, Lechner F;
 XX XX
 DR WPI; 2002-055561/07.
 XX XX
 PT New composition, useful for vaccine production, comprises antigen or
 PT antigenic determinant and non-natural molecular scaffold comprising
 PT organizer and core particle such as bacterial pilus or pilin protein
 PT -s
 PS
 PS Example 34; Page 265-267; 287pp; English.
 CC The invention relates to a composition comprising: (a) a non-natural
 CC molecular scaffold (molecular scaffold) which comprises a core
 CC particle such as a bacterial pilus or pilin protein, a recombinant form
 CC of the protein, a virus-like particle or a hepatitis B virus capsid
 CC protein (HBcAg), and an organizer; and (b) an antigen or antigenic
 CC determinant, where the molecular scaffold and antigenic determinant
 CC interact to form an ordered and repetitive antigen array. Suitable
 CC antigenic determinants include JON, FOS, HIV gp140, measles virus N
 CC protein, bee venom phospholipase, Sindbis virus E2 protein, amyloid beta
 CC derived peptides and influenza M2 antigen. The composition (or vaccine)
 CC is useful for immunisation, by administration to a subject, where the
 CC administration produces an immune response, such as humoral, cellular or
 CC protective immune response, preferably a Th type 2 T-helper (Th2)
 CC response that is specific for the antigenic determinant. The
 CC administration induces antibodies specific for the antigenic determinant
 CC of a subtype corresponding to the Th2 subtype in the subject. The subject
 CC does not generate a Th2 subtype that is specific for pilus or pilin
 CC polypeptide or antigenic determinant. The composition is useful for the
 CC production of vaccines for prevention of infectious diseases such as
 CC human immunodeficiency virus, viral hepatitis, measles, chicken pox,
 CC pneumonia, tuberculosis, syphilis, malaria, and for treating allergy,
 CC cancer, and chronic diseases induced or accelerated by a Th type
 CC immune response, such as arthritis, colitis, diabetes and multiple
 CC sclerosis. The composition is useful to generate defined self-specific
 CC antibodies and specific immune responses of the Th2 type and allows the
 CC creation of highly efficient vaccines against infectious diseases, and
 CC for treating allergy, cancer, and chronic diseases induced or accelerated

CC by a Th1 type immune response. The present segment is a nucleic acid
CC encoding a composition of the invention or a component of one.

Sequence 4637 BP; 1124 A; 1156 C; 1188 G; 1169 T; 0 other;

Alignment Scores:

Pred. No.:	2.08e-88	Length:	4637
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	24	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x ABK23921 (1-4637)

```
QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 3441 CCCGCGGATGATGGGCGCAAAACCTGCTGCTGATCTTTGACGCAAAATCTTTGGCAT 3500
QY 21 AsnAspTyProGluThrIleThrAspTyValThrLeuGlnArgGlySerAlaTyGly 40
DB 3501 AACGATTATCCGAAACCATTAACACTATGTACACTGCAACGAGGCTGGCTTATGGC 3560
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTySerGlySerTyProPhePro 60
DB 3561 GCGCTGTTATCTAATTTTCCGGACCGTAAATAATAGTGCCACTACCTATCCATTTCCT 3620
QY 61 ThrThrSerGluThrProArgValValTyAsnSerArgThrAspLysProThrProVal 80
DB 3621 ACCACGACGAAACGGCGCGCTGTTATATATTCGAAACGATTAAGCCGCGCGCTG 3680
QY 81 AlaLeuTyLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
DB 3681 GCGCTTATTTGACGCTGTGACAGTGCAGCGCGCTGCGCATTAAGCTGGCTCATTA 3740
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAspTyAsnSerAspPheGlnPheVal 120
DB 3741 ATTGCCGTGCTTATTTTGGCAGACACCAACACTATTAACGCGATGATTTCCAGTTTGTG 3800
QY 121 TrpAsnIleTyAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 3801 TGGATATTTACGCCAATATGATGTGTGCTGCTGCTGCGGCTGCATGTTTCGCT 3860
QY 141 ArgAspValThrValThrLeuProAspTyArgGlySerValProIleProLeuThrVal 160
DB 3861 CGGATGTCACCGTACTCTGCGGACTACCTGTTCACTGCAATTCCTTACCGTT 3920
QY 161 Tyr 161
DB 3921 TAT 3923
RESULT 14
AAD29374
ID AAD29374 standard; DNA; 7416 BP.
XX
AC AAD29374;
XX
DT 07-MAY-2002 (first entry)
XX
DE Plasmid pCGA126-1 for construction of plasmid pCGA139-1.
XX
KW FilmH: Immune response; antibacterial; enterobacillus-related disease;
KW therapy: vaccine; urinary tract infection; bladder; plasmid pCGA126-1;
KW ds.
XX
OS Chimeric - Escherichia coli.
OS Chimeric - unidentified.
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
PF 06-JUL-2001; 2001WO-US21525.
```

XX 07-JUL-2000; 2000US-216750P.

XX (MED1-) MEDIMUNE INC.

XX Langermann S, Revel A, Auguste C, Burteln J;

XX WPI; 2002-171702/22.

PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection

PS Example 2; Page 93-95; 101pp; English.

CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli film protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is plasmid pCGA126-1 used for constructing plasmid
CC pCGA139-1 used in the exemplification of the invention. Plasmid pCGA126-1
CC consists of Escherichia coli film DNA, filmC and its native signal and
CC lac p/o.

Sequence 7416 BP; 1858 A; 1807 C; 1908 G; 1843 T; 0 other;

Alignment Scores:

Pred. No.:	4.03e-88	Length:	7416
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	24	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29374 (1-7416)

```
QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 1244 CCCGCTGTAATGATGGGCGCAAAACCTGCTGCTGATCTTTGACGCAAAATCTTTGCCAT 1303
QY 21 AsnAspTyProGluThrIleThrAspTyValThrLeuGlnArgGlySerAlaTyGly 40
DB 1304 AACGATTATCCGAAACCATTAACACTATGTACACTGCAACGAGGCTGCTTATGGC 1363
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTySerGlySerTyProPhePro 60
DB 1364 GCGCTTATTTGACGCTGTGACAGTGCAGCGCGCTGCGCATTAAGCTGGCTCATTA 1423
QY 61 ThrThrSerGluThrProArgValValTyAsnSerArgThrAspLysProThrProVal 80
DB 1424 ACCACGACGAAACGGCGCGCTGTTATATTCGAAACGATTAAGCCGCGCGTG 1483
QY 81 AlaLeuTyLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
DB 1484 GCGCTTATTTGACGCTGTGACAGTGCAGCGCGCTGCGCATTAAGCTGGCTCATTA 1543
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAspTyAsnSerAspPheGlnPheVal 120
DB 1544 ATTGCCGTGCTTATTTTGGCAGACACCAACACTATTAACGAGGATTTCCAGTTTGTG 1603
QY 121 TrpAsnIleTyAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 1604 TGGATATTTACGCCAATATGATGTGTGCTGCTGCTGCGGCTGCAGTTTCCTCT 1663
QY 141 ArgAspValThrValThrLeuProAspTyArgGlySerValProIleProLeuThrVal 160
DB 1664 CGTATGTCACCGTACTCTGCGGACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1723
```



```

FT      /*tag- a
FT      /product= "E. coli FliM protein"
FT      /transl_except= (pos:526..528, aa:Arg)
FT      /transl_except= (pos:601..603, aa:His)
FT      /note= "CDS does not include start and stop codon"
FT      /partial
XX
XX
XX      WO200204496-A2.
XX
XX      17-JAN-2002.
XX
XX      06-JUL-2001; 2001WO-US21525.
XX
XX      07-JUL-2000; 2000US-216750P.
XX
XX      (MEDI-) MEDIMMUNE INC.
XX
XX      Langermann S, Revel A, Auguste C, Burteln J;
XX
XX      WPI; 2002-111702/22.
XX
XX      P-PDSB; AAE18435.
XX
XX      New immunogenic polypeptide, useful as vaccine for protecting against
XX      an enterobacillus-related disease in a patient at risk of contracting
XX      such disease, e.g. urinary tract infection or a bladder infection -
XX
XX      Example 1; Page 98; 101pp; English.
XX
XX      The invention relates to bacterial immunogenic agents for administration
XX      to humans and non-human animals to stimulate an immune response. The
XX      invention also relates to methods for vaccination of mammalian species
XX      with variants of E. coli FliM protein derived from different strains of
XX      E. coli. The vaccine composition or the antibody is useful for protecting
XX      against and treating an enterobacillus-related disease in a patient
XX      afflicted or at a risk of contracting the disease. In particular, the
XX      disease is a urinary tract or bladder infection. The present invention
XX      by a bacterium of the family Enterobacteriaceae, particularly E. coli.
XX      The present sequence is Escherichia coli strain G162 FliM DNA.
XX
XX      Sequence 837 BP; 192 A; 201 C; 221 G; 223 T; 0 other;
XX
XX
XX      Alignment Scores:
XX      Pred. No.: 4,25e-89 Length: 837
XX      Score: 831.00 Matches: 157
XX      Percent Similarity: 98.76% Conservative: 2
XX      Best Local Similarity: 97.52% Mismatches: 2
XX      Query Match: 98.00% Indels: 0
XX      DB: Gaps: 0
XX
XX      US-09-900-575-29_COPY_26_186 (1-161) x AAD29382 (1-837)
XX
XX      1 ProvalvalaslnvalglglnaslnleuvalvalaspleuSerhrGlnIlePheCySHS 20
XX      |||||
XX      Db 76 CCCGTCGTAATGTGGGGCAAAACCGTCGTCGATCTTCGACCGCAATCTTCCCAT 135
XX
XX      21 AasnApyrPrGluThrIleThrAspTyValThrLeuGlnArgIlySerAlaTyrgly 40
XX      |||||
XX      Db 136 AACGATTAATCCGGAACCACTTACAGACTATGTCACACTCAACAGACGCTCGGCTTAGGC 195
XX
XX      41 GIYAlleuSerAspPheSerGlyThrValLvsTySerGlySerSerTyrrProPhePro 60
XX      |||||
XX      Db 196 GCGCGTATATCAATTTTCCGGGACCGTAATAATTAAGGACAGTAAGCATTCATTCCT 255
XX
XX      61 ThrThrSerGluThrProArgValAlaTyraSerSerArgThrAspIlyProTrrProval 80
XX      |||||
XX      Db 256 ACCACACACGAAACCCGCGCTGTTATATTTGAGAGAACGATTAAGCCGCGCGG 315
XX
XX      81 AlaIeuTyrrIleuThrProvalSerSerAlaGlyIlyLeuValIleYsAlaGlySerIleu 100
XX      |||||
XX      Db 316 GCGCTTATTTGACCCCTGTGTGAGTAGTCGGGCGGGCGGCGCATTAAGCTGGTCATTA 375
XX
XX      101 IleAlaValleuIleLeuArgGlnThrAsnAsnTyraSerSerAspAspPheGlnIleVal 120
XX      |||||

```

Db	376	ATGCGCTGCTTATTTTGGCAGACACCAACAACATAACAGCAGATTTCCAGTTGTG	435
Oy	121	TrpAsnIleTyrIlaAsnAsnAspValIvalProthrgLyGlyCysAspValSerIa	140
Db	436	TGGAAATATTTACCCCAATATATGATGTGTGGTGCTACAGGGCGCTGGCATTTTCTGCT	495
Oy	141	ArgAspValThrValAlhIreupProAspIyFArgGlySerValProIleProLeuThrVal	160
Db	496	CGGAGATGTCACCGTATCTGTGCGGAGTACCTCGGTTCAGTGCCAAATTCCTTACGGTT	555
Oy	161	Tyr 161	
Db	556	TAT 558	
RESULT 17			
AAD29354	ID	AAD29354 standard; DNA; 837 BP.	
XX	AC	AAD29354;	
XX	AC		
DT	XX	07-MAY-2002 (first entry)	
DE	XX	Escherichia coli strain B217 F1mH DNA.	
KW	XX	F1mH; immune response; antibacterial; enterobacillus-related disease;	
XX	XX	therapy; vaccine; urinary tract infection; bladder; ds.	
OS	XX	Escherichia coli B217.	
FH	XX		
FT	FT	Key	
FT	FT	CDS	
FT	FT	Location/Qualifiers	
FT	FT	1..837	
FT	FT	/*tag= a	
FT	FT	/product= "E. coli F1mH protein"	
FT	FT	/transl_except= (pos:184..186, aa:Ala)	
FT	FT	/transl_except= (pos:208..210, aa:Ser)	
FT	FT	/transl_except= (pos:222..234, aa:Asn)	
FT	FT	/transl_except= (pos:326..328, aa:Arg)	
FT	FT	/transl_except= (pos:601..603, aa:His)	
FT	FT	/transl_except= (pos:700..702, aa:Ala)	
FT	FT	/transl_except= (pos:730..732, aa:Gly)	
FT	FT	/note= "CDS does not include start and stop codon"	
FT	FT	/partial	
PN	XX	WO200204496-A2.	
XX	XX		
PD	XX	17-JAN-2002.	
PF	XX	06-JUL-2001; 2001MO-US21525.	
XX	XX		
PR	XX	07-JUL-2000; 2000US-216750P.	
PA	XX	(MEDT-) MEDIMUNE INC.	
PI	XX	Langermann S, Revel A, Auguste C, Buirlein J;	
XX	XX	WPI: 2002-171702/22.	
DR	XX	P-PSDB; AAE18415.	
PT	XX	New immunogenic polypeptide, useful as vaccine for protecting against	
PT	XX	an enterobacillus-related disease in a patient at risk of contracting	
XX	XX	disease, e.g. urinary tract infection or a bladder infection	
PS	XX	Example 1: Fig 1; 101pp; English.	
CC	CC	The invention relates to bacterial immunogenic agents for administration	
CC	CC	to humans and non-human animals to stimulate an immune response. The	
CC	CC	invention also relates to methods for vaccination of mammalian species	
CC	CC	with variants of E. coli F1mH protein derived from different strains of	
CC	CC	E. coli. The vaccine composition or the antibody is useful for protecting	
CC	CC	against and treating an enterobacillus-related disease in a patient	
CC	CC	afflicted or at a risk of contracting the disease. In particular, the	
CC	CC	disease is a urinary tract or bladder infection. The disease is caused	
CC	CC	by a bacterium of the family Enterobacteriaceae, particularly E. coli.	

Db 619 TAT 621

RESULT 23
AA093068
ID AA093068 standard; DNA: 900 BP.

XX
AC AA093068;
XX
DT 15-MAR-1996 (first entry)
XX
DE FimH protein gene from E. coli clinical isolate MUJ31-3.
XX
KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW FimA; FimF; receptor binding site; ss.
XX
OS Escherichia coli clinical isolate MUJ31-3.

XX
FH Key Location/Qualifiers
FT sig_peptide 1..63
FT /*tag= a 64..900
FT mat_peptide /*tag= b

PN W09520657-A1.
XX
PD 03-AUG-1995.
XX
PF 27-JAN-1995; 95WO-DK00042.
XX
PR 27-JAN-1994; 94US-0187166.
XX
PA (GXBI-) GX BIOSYSTEMS AS.
XX
PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
PI MPI. 1995-275442/36.
DR P-PSDB: AAR76774.
XX
PT Receptor specific bacterial adhesins - useful for targetting active
PT compounds and microbial cells to locations of receptors
XX
PS Example 5; Page 69-74; 152pp; English.

XX The sequences given in AA093062-75 encode FimH proteins from various E.
CC coli clinical isolates. FimH is located at the tip of the type 1
CC fimbriae and also intercalated at intervals in the fimbrial organelle.
CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC structures containing terminally located alpha-D-mannoside residues.
CC FimH contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison FimA and
CC the minor components FimF and FimG only have 2 cysteine residues.
CC The localization of the cysteine residues in FimH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FimH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FimH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of variant FimH adhesins which
CC may be useful for targetting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.

XX
SQ Sequence 900 BP; 207 A; 209 C; 240 G; 244 T; 0 other;

Alignment Scores:
Pred. No.: 2.43e-88 Length: 900
Score: 825.00 Matches: 156
Percent Similarity: 97.52% Conservative: 1
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 97.29% Indels: 0
DB: 16 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AA093068 (1-900)

QY 1 ProValValAsnValGlyGlnAsnLeuValAlaAspLeuSerThrGlnIlePheCysHis 20
DB 139 CCHGCCGTGAATGTGGCGCAAAACCTGGTCGTAGATCTTTCGACGCAAACTCTTTGCCAT 198
QY 21 AsnAspTyrProGlnThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 199 AACGATTATCCGGAACACATTACAGACTATGTCACACTCGCAACGAGCTCGCTATGGC 258
QY 41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerTyrProPhePro 60
DB 259 GCGCTGTATCTAATTTTCCGGACCGTAAATATAGTGGCAGTAGCTATCCATTCCG 318
QY 61 ThrThrSerGlnThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
DB 319 ACTACACGCAAGCGCGGCTGTATTATATTCGAGAACGATTAAGCCGTGCCGGTG 378
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
DB 379 GCGCTTATTGTGACGCGCTGTAGACAGTGGCGGTGGCGCATTAAGCTGGCTCATTA 438
QY 101 IleAlaValLeuIleLeuAArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
DB 439 ATTCGCGCTTATTGTTCGACAGCACCACTATTAACGATGATTTCCAGTTGTG 498
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 499 TGGAAATATTACGCCAATATGATGTGTGTGCTGCTGCGGCTGCCATGTTTGTGCT 558
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 559 CATGATGTACACGTTACTCTCCGACTACCTGCTTCAGTGCATATTCCTTACCGTT 618
QY 161 Tyr 161
DB 619 TAT 621

RESULT 24
AAD29371
ID AAD29371 standard; DNA: 837 BP.

XX
AC AAD29371;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain G189 FimH DNA.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli G189.
XX
FH Key Location/Qualifiers
FT CDS 1..837
FT /*tag= a
FT /product= "E. coli FimH protein"
FT /transl_except= (pos:526..528, aa:Arg)
FT /transl_except= (pos:601..603, aa:His)
FT /note= "CDS does not include start and stop codon"
FT /partial

XX
PN W0200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MED1-) MEDIMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burtlein J;

XX WPI: 2002-171702/22.
DR P-PSDB: AAE18432.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
XX
XX Example 1; Page 66; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FliM protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain G189 FliM DNA.
XX
S0 Sequence 837 BP; 196 A; 200 C; 220 G; 221 T; 0 other:

Alignment Scores:
Pred. No.: 2.88e-88 Length: 837
Score: 824.00 Matches: 156
Percent Similarity: 97.52% Conservative: 1
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 97.17% Indels: 0
DB: Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29371 (1-837)

QY 1 ProvalValAsnValGlyGlnAsnLeuValAlaPleuSerThrGlnIlePheCysHis 20
Db CCTCCCGGATGTGGGGCAAAACCTGGTGTACATCTTTCGACGCCAAATCTTTGCCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
Db AACGATATTCGGAACCACTTACAGACTATGTCTACACTGCACAGCGCTGCGCTTATGCG 195
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
Db GGCCTGTTATCTATTTTCCGGGACCGGTAAATATAGTGGCACTATCCATTTCCG 255
QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
Db ACCACCGACGGAACCGCGCGGTGTTATTAATTCGAGAACGATAGCCGCGCGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
Db GCGCTTATTTGACCGCTGTGACGAGTGGCGGGGGGCGATTTAAAGCTGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
Db ATTGCCCTGCTTATTTTGGACAGACCAAAACTATAACGCGATATTTCCAGTTTG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
Db TGGATATTTTACCCCAATAATATGATGTGTAGTGTCTACATGCGCGCTGCTTCTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
Db CGTGATGTCAACGCTTACTGTGCGGACACTACCGCTGTTCACTGCTTCTTACCGTT 555
QY 161 Tyr 161
Db 556 TAT 558

RESULT 25
AAD29510
ID AAD29510 standard; DNA; 837 BP.
XX

AC AAD29510:
XX
XX 07-MAY-2002 (first entry)
XX
XX Escherichia coli strain EC189 FliM DNA.
DE
XX
XX FliM; immune response; antibacterial; enterobacillus-related disease;
KW vaccine; urinary tract infection; bladder; therapy; ds.
XX
XX Escherichia coli EC189.
OS
XX
XX WO200204496-A2.
PN
XX
XX 17-JAN-2002.
PD
XX
XX 06-JUL-2001; 2001WO-US21525.
PF
XX
XX 07-JUL-2000; 2000US-216750P.
PR
XX
XX (MEDT-) MEDIMUNE INC.
PA
XX
XX Langermann S, Revel A, Auguste C, Burlein J;
PI
XX
XX WPI: 2002-171702/22.
DR
XX
XX
XX New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
XX
XX Disclosure; Fig 1; 101pp; English.
XX
XX
XX The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FliM protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC189 FliM DNA.
XX
S0 Sequence 837 BP; 196 A; 200 C; 220 G; 221 T; 0 other:

Alignment Scores:
Pred. No.: 2.88e-88 Length: 837
Score: 824.00 Matches: 156
Percent Similarity: 97.52% Conservative: 1
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 97.17% Indels: 0
DB: Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29510 (1-837)

QY 1 ProvalValAsnValGlyGlnAsnLeuValAlaPleuSerThrGlnIlePheCysHis 20
Db CCTCCCGGATGTGGGGCAAAACCTGGTGTACATCTTTCGACGCCAAATCTTTGCCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
Db AACGATATTCGGAACCACTTACAGACTATGTCTACACTGCACAGCGCTGCGCTTATGCG 195
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
Db GGCCTGTTATCTATTTTCCGGGACCGGTAAATATAGTGGCACTATCCATTTCCG 255
QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
Db ACCACCGACGGAACCGCGCGGTGTTATTAATTCGAGAACGATAGCCGCGCGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
Db GCGCTTATTTGACCGCTGTGACGAGTGGCGGGGGGCGATTTAAAGCTGCTCATTA 375

```
OY 101 ILeAlaValleuIleuArgGIhThraSnaSPTyrAsnSerAspSphgInpheVal 120
Db 376 ATATCCCGTCTATTGTCGACAGACCAAACTATACAGCATGTATTCACACTTTGTG 435
OY 121 TrpAsnIleTyrAlaAsnAsnAspValValaValaProthrgIlyGlyCysAspValSerAla 140
Db 436 TCGAATATTATTCGCGCAATATGATGTAGTGCCTACTGCGGCTCGCATGTTTCTGCT 495
OY 141 ArgAspValThrValThrLeuProAspTyrArgIlySerValProIleProIleuThrVal 160
Db 496 CGTATGTACCGTACTCTGCGGACTACCGCTGTTCAAGTCCAAATTCCTTACCGTT 555
OY 161 Tyr 161
Db 556 TAT 558

RESULT 26
AAD29367
ID AAD29367 standard; DNA; 837 BP.
AC AAD29367;
XX 07-MAY-2002 (first entry)
DE Escherichia coli strain EC62 FimH DNA.
XX FimH; immune response; antibacterial; enterobacillus-related disease;
KM therapy; vaccine; urinary tract infection; bladder; ds.
XX Escherichia coli EC62.
OS
XX Key Location/Qualifiers
FH 1..837
FT CDS /product= "E. coli FimH protein"
FT /transl_except= (pos:526..528, aa:Arg)
FT /transl_except= (pos:601..603, aa:His)
FT /note= "CDS does not include start and stop codon"
XX
XX WO200204496-A2.
XX
XX 17-JAN-2002.
XX
XX 06-JUL-2001; 2001WO-US21525.
XX
XX 07-JUL-2000; 2000US-216750P.
XX
XX (MEDI-) MEDIMUNE INC.
XX
XX Langermann S, Revel A, Auguste C, Burlein J;
XX
XX WPI: 2002-171702/22.
XX
XX P-PSDB; AAEL8428.
XX
XX New immunogenic polypeptide, useful as vaccine for protecting against
XX an enterobacillus-related disease in a patient at risk of contracting
XX such disease, e.g. urinary tract infection or a bladder infection
XX
XX Example 1; Fig 1; 101pp; English.
XX
XX The invention relates to bacterial immunogenic agents for administration
XX to humans and non-human animals to stimulate an immune response. The
XX invention also relates to methods for vaccination of mammalian species
XX with variants of E. coli FimH protein derived from different strains of
XX E. coli. The vaccine composition or the antibody is useful for protecting
XX against and treating an enterobacillus-related disease in a patient
XX afflicted or at a risk of contracting the disease. In particular, the
XX disease is a urinary tract or bladder infection. The disease is caused
XX by a bacterium of the family Enterobacteriaceae, particularly E. coli.
XX The present sequence is Escherichia coli strain EC62 FimH DNA.
XX
```

```
SQ Sequence 837 BP; 195 A; 202 C; 221 G; 219 T; 0 other:
Alignment Scores:
Pred. No.: 4,99e-88 Length: 837
Score: 822.00 Matches: 155
Percent Similarity: 98.14% Conservative: 3
Best Local Similarity: 96.27% Mismatches: 3
Query Match: 96.93% Indels: 0
DB: 24 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29367 (1-837)
OY 1 ProValValaValAlGlyGlnAsnLeuValValaAspLeuSerThrGlnIlePheCysHis 20
Db 76 CCCGCCGTGAATGTGGGCAAAACCTGCTGCTAGATCTTCCAGCGAATCTTTGGCAT 135
OY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgIlySerAlaTyrGly 40
Db 136 AAGCATTTACCGGAAACCATTTACAGACTATGTCACACTGCACAGAGGTTGCGCTATGCG 195
OY 41 GlyValLeuSerAsnPheserGlyThrValIlySerGlySerSerTyrProPhePro 60
Db 196 GCGCTTATTTCATTTTCCGGGACCGTAAATATATAGTGCGATGCTATCCATTCTCT 255
OY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProThrProVal 80
Db 256 ACCACCGCAAGAACGCCGCCGCTGTTTATATTCAGAAACGATTAACCGTGGCGCGTG 315
OY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyIlyLeuValIleLysIlaGlySerLeu 100
Db 316 GCGCTTATTTCAGCGCTGTGACAGTGCAGTGGCGGTGGCATTAAGCTGGCTCATTA 375
OY 101 ILeAlaValleuIleuArgGIhThraSnaSPTyrAsnSerAspSphgInpheVal 120
Db 376 ATGCGCTGCTGAATTTTCCAGACAGACCAATACATATACAGCATGATTTCCAGTTGTG 435
OY 121 TrpAsnIleTyrAlaAsnAspValValaValaProthrgIlyGlyCysAspValSerAla 140
Db 436 TCGAATATTATTCGCGCAATATGATGTAGTGCCTGCGGCTCGCATGTTTCTGCT 495
OY 141 ArgAspValThrValThrLeuProAspTyrArgIlySerValProIleProIleuThrVal 160
Db 496 CGTATGTACCGTACTCTGCGGACTACCGCTGTTCAAGTCCAAATTCCTTACCGTT 555
OY 161 Tyr 161
Db 556 TAT 558

RESULT 27
AAD29369
ID AAD29369 standard; DNA; 837 BP.
AC AAD29369;
XX 07-MAY-2002 (first entry)
XX
XX Escherichia coli strain EC89 FimH DNA.
XX
XX FimH; immune response; antibacterial; enterobacillus-related disease;
XX therapy; vaccine; urinary tract infection; bladder; ds.
XX
XX Escherichia coli EC89.
XX
XX Key Location/Qualifiers
FH 1..837
FT CDS /product= "E. coli FimH protein"
FT /transl_except= (pos:526..528, aa:Arg)
FT /transl_except= (pos:601..603, aa:His)
FT /transl_except= (pos:676..678, aa:Thr)
FT /transl_except= (pos:679..681, aa:Ala)
FT /transl_except= (pos:691..693, aa:Ile)
FT /transl_except= (pos:694..696, aa:Val)
```

QY 121 TTPASNIETyrAlaAsnAsnAspValValProthrGlyCysaspValSerAla 140

may be useful for targeting active compounds and microbial cells to locations comprising selected receptors to which the adhesion binds

Sequence 900 BP; 206 A; 214 C; 241 G; 239 T; 0 other,


```

DB 496 CATGATGTCACCGTACTGCGGACATACCTGGTTCAGTCCCAATTCCTTACCGTT 555
QY 161 Tyr 161
DB 556 TAT 558
RESULT 30
ID AAD29362 standard; DNA; 837 BP.
XX
AC AAD29362;
XX
DE 07-MAY-2002 (first entry)
DE Escherichia coli strain EC45 FimH DNA.
XX
KM FimH; immune response; antibacterial; enterobacillus-related disease;
XX therapy; vaccine; urinary tract infection; bladder; ds.
OS Escherichia coli EC45.
XX
FH Key Location/Qualifiers
FT CDS 1..837
FT /tag= a
FT /product= "E. coli FimH protein"
FT /note= "CDS does not include start and stop codon"
XX
PN MO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burtlein J;
XX
DR WPI; 2002-171702/22.
XX
DR P-PSDB; AAE18423.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
XX
PS Example 1; Fig 1; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC45 FimH DNA.
XX
SQ Sequence 837 BP; 193 A; 202 C; 224 G; 218 T; 0 other;
XX
Alignment Scores:
Pred. No.: 1,136-87 Length: 837
Score: 819.00 Matches: 154
Percent Similarity: 98.14% Conservative: 4
Best Local Similarity: 95.65% Mismatches: 3
Query Match: 96.58% Indels: 0
DB: 24 Gaps: 0
US-09-900-575-29_COPY_26_186 (1-161) x AAD29362 (1-837)

```

```

QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 76 CTGCGCGTGAATGTTGGGCAAAACCTGCTGCTGATCTTTCAGACGCAAAATCTTCCCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 136 AAGCATTTACCCAGAAACCATTTACAGACTATGTACACCTGCACACGAGGTGGCTTATGGC 195
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
DB 196 GGGCTTTATCTGATCTTTTCCGGACCGTTAAATATATATGCGAGTATGCTATCTTCCCT 255
QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
DB 256 ACTACACGCGAAACGCCGCGGTTGTTTAAATTCAGAAACGATTAAGCCGTCGCGGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleValLysAlaGlySerLeu 100
DB 316 GCGCTTTATTTGACGCGCGGACAGTCCGGGGGAGTGCGCATTTAAAGCTGCTTATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
DB 376 ATTGCCGCTGCTTATTTTGGACAGACCAACTATTAACAGGATTTCCAGTTTGG 435
QY 121 TrpAsnIleTyrAlaAsnAspValValAlaProThrGlyGlyCysAspValSerAla 140
DB 436 TGGAAATATTTAGCCCATATATGATGTGTGCCACACGCGGCGGCTCATTTCTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 496 CGTGATGTCACCGTACTGCTGCGGACCTACCTCGTTGATGCTCCATTTACCGTT 555
QY 161 Tyr 161
DB 556 TAT 558
RESULT 31
ID AAD29373
XX
AC AAD29373 standard; DNA; 837 BP.
XX
AC AAD29373;
XX
DE 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain NU14 FimH DNA.
XX
KM FimH; immune response; antibacterial; enterobacillus-related disease;
XX therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli NU14.
XX
FH Key Location/Qualifiers
FT CDS 1..837
FT /tag= a
FT /product= "E. coli FimH protein"
FT /transl_except= (pos:526..528, aa:Arg)
FT /transl_except= (pos:601..603, aa:His)
FT /note= "CDS does not include start and stop codon"
XX
PN MO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burtlein J;
XX
DR WPI; 2002-171702/22.

```

DR P-PSDB; AAE18434.

XX New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
XX
XX
PS Example 1; Page 67; 101pp; English.

XX The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain N014 FimH DNA.

SQ Sequence 837 BP; 193 A; 202 C; 224 G; 218 T; 0 other;

Alignment Scores:

Pred. No.:	1,13e-87	Length:	837
Score:	819.00	Matches:	154
Percent Similarity:	98.14%	Conservative:	4
Best local Similarity:	95.65%	Mismatches:	3
Query Match:	96.58%	Indels:	0
DB:	24	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AMD29373 (1-837)

QY 1 ProValValAsnValGlyGlnAsnLeuValAlaSplLeuSerThnGlnIlePheCysHis 20
DB 76 CCGCGCGAATGTGGCGCAAAACCTGCTGAGATCTTTGAGCCCAATCTTTGGCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 136 AACGATACCCAGAACCATTTACAGACTATGTCACACGCAAGCGGTGCGCTATGCGC 195
QY 41 GlyValLeuSerAsnPheSerGlyThrValIlySyrSerGlySerTyrProPhePro 60
DB 196 GCGCGTATCTAGTTTTCGCGGACCCGTAATAATATGAGTATGCTATCCCTTCCT 255
QY 61 ThrThrSerGluThrProArgValAlaTyrAsnSerArgThrAspLysProThrProVal 80
DB 256 ACTACAGGAAAGCGCGCGGTGTTTATTAATTCGAGAACGATTAAGCCGTGGCGGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
DB 316 GCGCTTATTGACGCGCGGTGAGCAGTGGGGGAGTGGCGATTAAAGCTGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
DB 376 ATGCGCGTCTATTTTGGACAGACCAACTATTAACAGGAGATTCACAGTTTGTG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 436 TGGAAATATTAGCCCAATATGATGTGTGTGTCACACTGGCGGTGCGCATTTCTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 496 CGTGATGTACCGTACTCTGCGCGACTACCGTTCAGTTCGCGATTCCTTACCGTT 555
QY 161 Tyr 161
DB 556 TAT 558

RESULT 32

AMD29509

ID AMD29509 standard; DNA; 837 BP.

XX AAD29509;

XX

DT 07-MAY-2002 (first entry)

XX Escherichia coli strain B203 FimH DNA.

XX FimH; immune response; antibacterial; enterobacillus-related disease;
KW vaccine; urinary tract infection; bladder; therapy; ds.
XX
XX
OS Escherichia coli B203.

XX WO200204496-A2.

XX 17-JAN-2002.

XX 06-JUL-2001; 2001WO-US21525.

XX 07-JUL-2000; 2000US-216750P.

XX (MEDI-) MEDIMUNE INC.

XX Langermann S, Revel A, Auguste C, Burtin J;

XX WPI; 2002-171702/22.

XX New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
XX
PS Disclosure; Fig 1; 101pp; English.

XX The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain B203 FimH DNA.

SQ Sequence 837 BP; 193 A; 201 C; 225 G; 218 T; 0 other;

Alignment Scores:

Pred. No.:	1,13e-87	Length:	837
Score:	819.00	Matches:	154
Percent Similarity:	98.14%	Conservative:	4
Best local Similarity:	95.65%	Mismatches:	3
Query Match:	96.58%	Indels:	0
DB:	24	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AMD29509 (1-837)

QY 1 ProValValAsnValGlyGlnAsnLeuValAlaSplLeuSerThnGlnIlePheCysHis 20
DB 76 CCGCGCGAATGTGGCGCAAAACCTGCTGAGATCTTTGAGCCCAATCTTTGGCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 136 AACGATACCCAGAACCATTTACAGACTATGTCACACGCAAGCGGTGCGCTTATGCGC 195
QY 41 GlyValLeuSerAsnPheSerGlyThrValIlySyrSerGlySerTyrProPhePro 60
DB 196 GCGCGTATCTAGTTTTCGCGGACCCGTAATAATATGAGTATGCTATCCCTTCCT 255
QY 61 ThrThrSerGluThrProArgValAlaTyrAsnSerArgThrAspLysProThrProVal 80
DB 256 ACTACAGGAAAGCGCGCGGTGTTTATTAATTCGAGAACGATTAAGCCGTGGCGGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
DB 316 GCGCTTATTGACCGCGGTGAGCAGTGGGGGAGTGGCGATTAAAGCTGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120


```

|||||
Db 376 ATGGCGTCTATTGTCGACAGACCAACTATTAACAGGATGTTCCAGTTTGG 435
Oy 121 TTPASNIETYLALASNASNPVALVALVALPROTHGLYCYSAPVALSERALA 140
Db 436 TGGAAATATTACCCCAATATGATGTGTGGCCACGCGCGCTGCATGTTCTGCT 495
Oy 141 ARGASPVALTHTVALTHRLEUPROASPTYRARGGLYSERVALPROLEPROLEUTHRAL 160
Db 496 CGTGATGTCACCGTTACTCTGCGGACATACCTGGTTCACTGCCATTCCTTACCGTT 555
Oy 161 TYR 161
Db 556 TAT 558

RESULT 33
AAD29355
ID AAD29355 standard; DNA: 840 BP.
XX
AC AAD29355;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain B223 F1mH DNA.
XX
KW F1mH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli B223.
XX
FH
FT Key location/Qualifiers
FT CDS 1..840
FT /tag- a
FT /product- "E. coli F1mH protein"
FT /transl_except- (pos:526..528, aa:Arg)
FT /transl_except- (pos:601..603, aa:His)
FT /note- "CDS does not include start codon"
XX
PN W0200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001MO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDT-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burteln J;
XX
DR WPI: 2002-117102/22.
XX
DR P-PSDB; AAE18416.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
XX
PS Example 1; Page 59; 101bp; English.
XX
XX The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli F1mH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain B223 F1mH DNA.
XX
SO Sequence 840 BP; 195 A; 202 C; 224 G; 219 T; 0 other;

```

```

Alignment Scores:
Pred. No.: 1.14e-87 Length: 840
Score: 819.00 Matches: 154
Percent Similarity: 98.14% Conservative: 4
Best Local Similarity: 95.65% Mismatches: 3
Query Match: 96.58% Indels: 0
DB: 24 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29355 (1-840)
Oy 1 ProValValAsnValGlyGlnAsnLeuValAlaAspLeuSerThrGlnIlePheCysHis 20
Db 76 CTGCGCGTGAATGGGGCAAAACCTGGTCGTGATCTTTCACGCAAAATCTTCCAT 135
Oy 21 AsnAspTyrProGluPrlIleThrAspTyrValThrluGlnArgGlySerAlaTyrGly 40
Db 136 AAGGATTACCCAGAAACCATTAACAGACTATGTCACATCGAACGAGAGTGGCCTATGGC 195
Oy 41 GlyValLeuSerAsnDheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
Db 196 GGCCTGTATCTAGTTTTCGCGGACCGTAATATATATGAGAGTATCTATCTTCCCT 255
Oy 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrrProVal 80
Db 256 ACTACACAGCAAAACCGCGGGTGTGTATTAATCGAAGAACGATTAACCGTGGCGGTG 315
Oy 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
Db 316 GCGCTTATTTGACGCGCGGTGAGACGTGCGGGGAGTGCCATTAACCTGGCTCATTTA 375
Oy 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
Db 376 ATGGCGTCTATTGTCGACAGACCAACATTAACAGCAGATGATTTCCAGTTTGG 435
Oy 121 TTPASNIETYLALASNASNPVALVALVALPROTHGLYCYSAPVALSERALA 140
Db 436 TGGAAATATTACCCCAATATGATGTGTGGCCACGCGCGCTGCATGTTCTGCT 495
Oy 141 ARGASPVALTHTVALTHRLEUPROASPTYRARGGLYSERVALPROLEPROLEUTHRAL 160
Db 496 CGTGATGTCACCGTTACTCTGCGGACATACCTGGTTCACTGCCATTCCTTACCGTT 555
Oy 161 TYR 161
Db 556 TAT 558

RESULT 34
AAD29360
ID AAD29360 standard; DNA: 837 BP.
XX
AC AAD29360;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain DS17 F1mH DNA.
XX
KW F1mH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli DS17.
XX
FH
FT Key location/Qualifiers
FT CDS 1..837
FT /tag- a
FT /product- "E. coli F1mH protein"
FT /transl_except- (pos:526..528, aa:Arg)
FT /transl_except- (pos:601..603, aa:His)
FT /note- "CDS does not include start and stop codon"
XX
PN W0200204496-A2.
PD 17-JAN-2002.

```

XX	06-JUL-2001; 2001WO-US21525.
XX	07-JUL-2000; 2000US-216750P.
XX	(MED1-) MEDIMMUNE INC.
XX	
XX	Langermann S, Revel A, Auguste C, Burteln J.
XX	WPI: 2002-171702/22.
XX	P-PSDB: AAE18421.
XX	
XX	New immunogenic polypeptide, useful as vaccine for protecting against
XX	an enterobacillus-related disease in a patient at risk of contracting
XX	such disease, e.g. urinary tract infection or a bladder infection
XX	
XX	Example 1; Fig 1; 101pp; English.
XX	
XX	The invention relates to bacterial immunogenic agents for administration
XX	to humans and non-human animals to stimulate an immune response. The
XX	invention also relates to methods for vaccination of mammalian species
XX	with variants of E. coli FimH protein derived from different strains of
XX	E. coli. The vaccine composition or the antibody is useful for protecting
XX	CC against and treating an enterobacillus-related disease in a patient
XX	CC afflicted or at a risk of contracting the disease. In particular, the
XX	CC disease is a urinary tract or bladder infection. The disease is caused
XX	by a bacterium of the family Enterobacteriaceae, particularly E. coli.
XX	The present sequence is Escherichia coli strain DS17 FimH DNA.
XX	
XX	Sequence 837 BP; 194 A; 200 C; 223 G; 220 T; 0 other;
XX	
XX	Alignment Scores:
XX	Pred. No.: 1,49e-87 Length: 837
XX	Score: 818.00 Matches: 154
XX	Percent Similarity: 97.52% Conservative: 3
XX	Best Local Similarity: 95.65% Mismatches: 4
XX	Query Match: 96.46% Indels: 0
XX	DB: 24 Gaps: 0
XX	
XX	US-09-900-575-29_COPY_26_186 (1-161) x AAD29360 (1-837)
XX	
XX	1 proValValAspValGlyClnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
XX	76 CCGCCGCGTAATGTGGGCAAACTGTCGTAGATCTTTCGACGCCAATCTTTGCCAT 135
XX	
XX	21 AsnAspTYrProGluThrIleThrAspTYrValThrLeuGlnAArgIySerAlaTYrGly 40
XX	136 AACGATATCCAGAAACATTACAGACTATGTCACTCGACAGAGGTTGCGCTTATGGC 195
XX	
XX	41 GlyValLeuSerAspPheSerGlyThrValLysTYrSerGlySerSerTYrProPhePro 60
XX	196 GGGGTGTATCTAGATTTTCGGGACCTTAATAATATGACAGTACTACTTTCCTTCCCT 255
XX	
XX	61 ThrThrSerGluThrProArgValValTYrAsnSerArgThrAspLysProThrProVal 80
XX	256 ACTACACGGGAAGCGCGGCTTTTATTAATTCGAGACGGAATAGCCGTGGCGGTG 315
XX	
XX	81 AlaLeuTYrLeuThrProValSerSerIleGlyGlyLeuValIleLysAlaGlySerLeu 100
XX	316 GCGCTTATATTTGACCGCGGTGAGTCAGTCGGGGGAGTGGCGATTAAGCTGGCTATT 375
XX	
XX	101 IleAlaValLeuIleLeuArgGluThrAsnAsnTYrAsnSerAspAspPheGlnPheVal 120
XX	376 ATTGCCGTGCTTATTTTGGACAGACCAACAACATATACACGATGATTTCCAGTTTGTG 435
XX	
XX	121 TrpAsnIleTYrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
XX	436 TGGAAATATTACCCCAATATGATGTGGGTGGCCCACTGGCGGCTGTGATGCTTCTCT 495
XX	
XX	141 ArgAspValThrValThrLeuProAspTYrArgGlySerValProIleProLeuThrVal 160
XX	496 CGGATGTACCGCTTACTTTGGCGAGACTACCTGGTTCAGTGGCGAATTCCTTACCGTT 555

```

Oy      161 Tyr 161
        111
Db      556 TAT 558

RESULT 35
AAD29352
ID      AAD29352 standard: DNA; 837 BP.
XX
XX      AAD29352;
XX
XX      07-MAY-2002 (first entry)
XX
XX      Escherichia coli strain B210 FimH DNA.
XX
XX      FimH; immune response; antibacterial; enterobacillus-related disease;
KW      therapy; vaccine; urinary tract infection; bladder; ds.
XX
XX      Escherichia coli B210.
XX
XX      Key      Location/Qualifiers
XX      CDS      1..837
XX                /tag= a
XX                /product= "E. coli FimH protein"
XX                /transl_except= (pos:421..423, aa:Arg)
XX                /transl_except= (pos:526..528, aa:Arg)
XX                /transl_except= (pos:601..603, aa:His)
XX                /transl_except= (pos:607..609, aa:Asp)
XX                /note= "CDS does not include start and stop codon"
XX                /partial
XX
XX      WO200204496-A2.
XX
XX      17-JAN-2002.
XX
XX      06-JUL-2001; 2001WO-US21525.
XX
XX      07-JUL-2000; 2000US-216750P.
XX
XX      (MEDI-) MEDIMUNE INC.
XX
XX      Langermann S, Revel A, Auguste C, Burslein J;
XX
XX      WPI: 2002-1171702/22.
XX      P-PSDB; AAE18413.
XX
XX      New immunogenic polypeptide, useful as vaccine for protecting against
XX      an enterobacillus-related disease in a patient at risk of contracting
XX      such disease, e.g. urinary tract infection or a bladder infection -
XX
XX      PS      Example 1; Fig 1; 101pp; English.
XX
XX      The invention relates to bacterial immunogenic agents for administration
XX      to humans and non-human animals to stimulate an immune response. The
XX      invention also relates to methods for vaccination of mammalian species
XX      with variants of E. coli FimH protein derived from different strains of
XX      E. coli. The vaccine composition or the antibody is useful for protecting
XX      against and treating an enterobacillus-related disease in a patient
XX      afflicted or at a risk of contracting the disease. In particular, the
XX      disease is a urinary tract or bladder infection. The disease is caused
XX      by a bacterium of the Family Enterobacteriaceae, particularly E. coli.
XX      The present sequence is Escherichia coli strain B210 FimH DNA.
XX
XX      Sequence 837 BP; 166 A; 204 C; 228 G; 219 T; 0 other;
XX
XX      Alignment Scores:
XX      Pred. No.:      1..96e-87      Length:      837
XX      Score:          817..00      Matches:      155
XX      Percent Similarity: 96.89%      Conservative: 1
XX      Best Local Similarity: 96.27%      Mismatches:   5
XX      Query Match:      96.34%      Indels:      0
XX      Db:              24      Gaps:          0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29352 (1-837)

```

```

QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 76 CCGCTGCTGAATGTGGGCAAAACCTGCTGATCTTTCACCAAACTTTTGGCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 136 AAGCATTTACCGAAGCAATACAGACTATGTACACACTGCACAGAGCTCGGCTTATGCG 195
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
DB 196 GCGGTGTATCTTAATTTTCCGGATCGTAATAATATAGTGCAGATCTATCTTTCCCT 255
QY 61 ThrThrSerGluThrProArgValValLysAsnSerArgThrAspLysProTyrProVal 80
DB 256 ACCACACGCAAGCAAGCCGCGCTGTATATTCAGACAGATTAAGCCGTGCGCGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
DB 316 GCGCTTTATTTGACCCCTGTGACAGCTCCGGGGAGTGGCGATTAAGCGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
DB 376 ATTGCCGTCTTATTTTCCGACAGCAACAATAAACCAGATGCTTCCAGTTTGTG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 436 TGGAAATATTTACGCCAATATGATGTGGTCCCACTGCGCGCTCGATGCTTGTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgLysSerValProIleProLeuThrVal 160
DB 496 CGTGATGTACACGTTACTGCTCCGAGTACCCTGTCTCATGTGCGGATTCCTTACCGTT 555
QY 161 Tyr 161
DB 556 TAT 558

RESULT 36
AAD29364
ID AAD29364 standard; DNA: 837 BP.
XX
AC AAD29364:
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC58 film DNA.
XX
KW film: immune response; antibacterial; enterobacillus-related disease;
XX therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli EC58.
XX
FH key Location/Qualifiers
FT CDS 1..837
FT /tag= a
FT /product= "E. coli film protein"
FT /transl_except= (pos:526..528, aa:Arg)
FT /transl_except= (pos:601..603, aa:His)
FT /note= "CDS does not include start and stop codon"
FT /partial
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001MO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burteln J,
XX

```

```

DR WPI; 2002-171702/22.
DR P-PSDB: AAE18425.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
XX an enterobacillus-related disease in a patient at risk of contracting
XX such disease, e.g. urinary tract infection or a bladder infection
PS Example 1; Fig 1; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli film protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC58 film DNA.
XX
SO Sequence 837 BP; 195 A; 200 C; 222 G; 220 T; 0 other;

Alignment Scores:
Pred. No.: 2,586-87 Length: 837
Score: 816.00 Matches: 154
Percent Similarity: 97.52% Conservative: 3
Best Local Similarity: 95.65% Mismatches: 4
Query Match: 96.23% Indels: 0
DB: 24 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29364 (1-837)
QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 76 CCGCTGCTGAATGTGGGCAAAACCTGCTGATCTTTCACCAAACTTTTGGCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 136 AAGCATTTACCGAAGCAATACAGACTATGTACACACTGCACAGAGTTCGCGCTTATGCGC 195
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
DB 196 ACGGTATATCTAGTTTTCGCGACCGTAATAATATATAGCATCTATCTTTCCT 255
QY 61 ThrThrSerGluThrProArgValValLysAsnSerArgThrAspLysProTyrProVal 80
DB 256 ACTACACGCAAGCAAGCCGCGGCTGTATTAATTCGAGAACGGATTAAGCCGTGCGCGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
DB 316 GCGCTTTATTTGACGCGCGGTGAGCAGTGCAGGAGATGCGCATTAAGCTGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
DB 376 ATTGCCGTCTTATTTTCCGACAGCAACAATAAACCAGATGCTTCCAGTTTGTGCT 495
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 436 TGGAAATATTTACGCCAATATGATGTGGTCCCACTGCGCGCTGTGCTTGTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgLysSerValProIleProLeuThrVal 160
DB 496 CGTGATGTACACGTTACTGCTCCGAGTACCCTGTCTCATGTGCGGATTCCTTACCGTT 555
QY 161 Tyr 161
DB 556 TAT 558

RESULT 37
AAD29365
ID AAD29365 standard; DNA: 837 BP.
XX
AC AAD29365:
XX

```

```

XX 07-MAY-2002 (first entry)
DT Escherichia coli strain EC60 FimH DNA.
DE
XX FimH: immune response; antibacterial; enterobacillus-related disease;
KM therapy; vaccine; urinary tract infection; bladder; ds.
XX Escherichia coli EC60.
OS
XX Key
FH Location/Qualifiers
FT CDS
FT 1..837
FT /product= "E. coli FimH protein"
FT /transl_except= (pos:79..81, aa:Val)
FT /transl_except= (pos:196..198, aa:Gly)
FT /transl_except= (pos:208..210, aa:Asn)
FT /transl_except= (pos:232..234, aa:Ser)
FT /transl_except= (pos:355..357, aa:Val)
FT /transl_except= (pos:526..528, aa:Arg)
FT /transl_except= (pos:601..603, aa:His)
FT /transl_except= (pos:805..807, aa:Arg)
FT /transl_except= (pos:817..819, aa:Ala)
FT /note= "CDS does not include start and stop codon"
FT /partial
XX
XX MO200204496-A2.
XX
XX 17-JAN-2002.
XX
XX 06-JUL-2001; 2001WO-US21525.
XX
XX 07-JUL-2000; 2000US-216750P.
XX
XX (MED-) MEDIMMUNE INC.
XX
XX Langermann S, Revel A, Auguste C, Buirlein J;
XX
XX WPI; 2002-171702/22.
XX
XX P-PSDB; ABE18426.
XX
XX New immunogenic polypeptide, useful as vaccine for protecting against
XX an enterobacillus-related disease in a patient at risk of contracting
XX such disease, e.g. urinary tract infection or a bladder infection
XX
XX Example 1; Fig 1; 101pp; English.
XX
XX The invention relates to bacterial immunogenic agents for administration
XX to humans and non-human animals to stimulate an immune response. The
XX invention also relates to methods for vaccination of mammalian species
XX with variants of E. coli FimH protein derived from different strains of
XX E. coli. The vaccine composition or the antibody is useful for protecting
XX against and treating an enterobacillus-related disease. In particular, the
XX afflicted or at a risk of contracting the disease. In particular, the
XX disease is a urinary tract or bladder infection. The disease is caused
XX by a bacterium of the family Enterobacteriaceae, particularly E. coli.
XX The present sequence is Escherichia coli strain EC60 FimH DNA.
XX
XX Sequence 837 BP; 195 A; 200 C; 222 G; 220 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 2 58e-87 Length: 837
XX Score: 816.00 Matches: 154
XX Percent Similarity: 97.52% Conservative: 3
XX Best Local Similarity: 95.65% Mismatches: 4
XX Query Match: 96.23% Indels: 0
XX DB: 24 Gaps: 0
XX
XX US-09-900-575-29_COPY_26_186 (1-161) x AMD29365 (1-837)
XX
XX 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
XX 76 CCTGCGTGAATGTGGGCAAAACCTGCTAGATCTTTCGACGCAAAATCTTTGGCAT 135

```

```

XX 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
XX 136 AACGATTACCCAGAAACCACTTACAGACTATGTCACACTGCACAGGATTGCGCTTATGCG 195
XX 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyr-ProPhePro 60
XX 196 ACGGTTTATCTAGTTTTCGCGGACCGTAAATATATATGAGAGTATCTTCCT 255
XX 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProThrProVal 80
XX 256 ACTACACAGCAACGCCGCGGCTGTTATATTCGAGAACGATAACCCGTCGCGGCTG 315
XX 81 AlaLeuTyrLeuThrProValSerSerIleGlyGlyLeuValIleLysAlaGlySerLeu 100
XX 316 GCGCTTATTGTCGCGGCTGAGCACTGCGGCGGAGTGCACATTAAAGTCGTCANTTA 375
XX 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
XX 376 ATTGCCGTGCTTATTTTGGACAGACCAACACTATACAGCAGATTTCCAGTTTG 435
XX 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
XX 436 TGAATATTACGCCAATATGATGTGTGTCGCCACTGCGGCGCTGATGTTCTGCT 495
XX 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
XX 496 CGTGATGTCACCGTTACTGTCGCGGACTACCCGTCGATGAGTCGATTCCTTACCGTT 555
XX 161 Tyr 161
XX 556 TAT 558
XX
XX RESULT 38
XX ABL52756
XX ID ABL52756 standard; DNA; 1227 BP.
XX ABL52756;
XX
XX 01-JUL-2002 (first entry)
XX
XX Fusion protein prty-FimH-prty encoding sequence.
XX
XX Exoprotease; immunostimulant; vaccine; anchor peptide; gene;
XX FimH; ciliated adhesive factor; ds.
XX
XX Lactobacillus helveticus.
XX Escherichia coli.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 1..1227
XX /tag= a
XX /product= "Prty-FimH-Prty"
XX
XX JP2002017357-A.
XX
XX 22-JAN-2002.
XX
XX 04-JUL-2000; 2000JP-0202442.
XX
XX 04-JUL-2000; 2000JP-0202442.
XX
XX (CALV ) CALPIS SHOKUJIN KOGYO KK.
XX
XX WPI; 2002-221706/28.
XX P-PSDB; ABB09458.
XX
XX New anchor peptide, useful for anchoring protein onto microbe
XX
XX Example 1; Page 8-10; 12pp; Japanese.
XX
XX The invention relates to an anchor peptide derived from the

```

CC exoproteinase of *Lactobacillus helveticus* FERM BP-6060.
 CC The activity of compositions of the invention may be described as
 CC immunostimulatory. The invention also includes a fusion protein,
 CC containing the anchor peptide and a signal peptide fused to a useful
 CC protein. The invention also includes a method for annotating a useful
 CC protein onto the surface of a microbe. The microbe can be used as a
 CC vaccine. The current sequence represents an encoding sequence for a
 CC fusion protein referred to as pFty-FimH-prty. This fusion protein
 CC contains a fragment of the FimH protein, which is a ciliated adhesive
 CC factor from *E. coli*, arranged between the signal sequence and anchor
 CC sequences derived from *L. helveticus* exoproteinase (see ABB09456 and
 CC ABB09457).
 CC
 XX Sequence 1227 BP; 325 A; 285 C; 287 G; 330 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 4 41e-87 Length: 1227
 Score: 816.00 Matches: 154
 Percent Similarity: 97.52% Conservative: 3
 Best Local Similarity: 95.65% Mismatches: 4
 Query Match: 96.23% Indels: 0
 DB: 24 Gaps: 0
 US-09-900-575-29_COPY_26_186 (1-161) x ABL52756 (1-1227)
 QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
 DB 277 CCTCCGCGTGAATGTGGGCGCAAAAGCTGTGCTAGATCTTTGACGCAATCTTTGGCAT 336
 QY 21 AsnAspPyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
 DB 337 AACGATTACCCAGAAACCATTAACACTATGTACACTGCACACGAGGTGGCGTTATGGC 396
 QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
 DB 397 GCGGTGTATCTAGCTTTTCCGGGACCGTAAATATATGACACTATCTTTCCT 456
 QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
 DB 457 ACTACCGACGAAACCGCGGCTGTTTATTAATTCGAGAACGGATAGCCGTCGCGTG 516
 QY 81 AlaleuThrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
 DB 517 GCCGTTATTTGACCGCGGTGAGAGTGGCGGAGTGCAGTAAAGCTGCTCATTTA 576
 QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAspTyrAsnSerAspPheGlnPheVal 120
 DB 577 ATTGCCGTGCTTATTTTGGCAGACCAACACTATACAGCGATGTTCCAGTTTGTG 636
 QY 121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyLysAspValSerAla 140
 DB 637 TGGAAATATTACCAATATATGATGTGGTCCGCTGCGGCGTGCATGTTTGCCT 696
 QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
 DB 697 CGTGATGTCCACCGTTACTGTGCGGACTACCTGTTCACTGCGGATTCCTTACCGTT 756
 QY 161 Tyr 161
 DB 757 TAT 759
 RESULT 39
 AAQ93066
 ID AAQ93066 standard; DNA: 900 BP.
 XX
 AC AAQ93066;
 XX
 DT 15-MAR-1996 (first entry)
 XX
 DE FimH protein gene from *E. coli* clinical isolate MJ#2-2.
 XX
 KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
 FimA; FimF; FimG; receptor binding site; ss.

XX
 OS *Escherichia coli* clinical isolate MJ#2-2.
 XX
 FH Key Location/Qualifiers
 FT sig_peptide 1..63
 FT mat_peptide /*lag- a 64..900
 FT /*lag- b
 XX
 PN W09520657-A1.
 XX
 PD 03-AUG-1995.
 XX
 PE 27-JAN-1995; 95MO-DK00042.
 XX
 PR 27-JAN-1994; 94US-0187166.
 XX
 PA (GXBI-) GX BIOSYSTEMS AS.
 XX
 PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
 XX
 DR WPI; 1995-275442/36.
 XX
 PT P-PSDB; AAR76776.
 XX
 PS Receptor specific bacterial adhesins - useful for targetting active
 PT compounds and microbial cells to locations of receptors
 XX
 PS Example 5; Page 69-74; 152pp; English.
 XX
 CC The sequences given in AAQ93062-75 encode FimH proteins from various *E.*
 CC *coli* clinical isolates. FimH is located at the tip of the type 1
 CC fimbriae and also intercalated at intervals in the fimbrial organelle.
 CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
 CC structures containing terminally located alpha-D-mannoside residues.
 CC FimH contains 4 cysteine residues assumed to direct folding of the
 CC molecule into distinct functional domains. For comparison FimA and
 CC the minor components FimF and FimG only have 2 cysteine residues.
 CC The localisation of the cysteine residues in FimH points to a tandem
 CC arrangement of two ancestral genes. Similar amino acids can be
 CC found in similar positions in the two halves of the FimH protein. The
 CC 'midway' point is located roughly around residue 150 in the mature
 CC protein. The two halves or domains of FimH have evolved differently
 CC with the N-terminal section becoming the domain harbouring the receptor
 CC binding site, whereas the C-terminal sector became the domain of the
 CC molecule required for integration into the fimbrial organelle. These
 CC sequences may be used in the production of variant FimH adhesins which
 CC may be useful for targetting active compounds and microbial cells to
 CC locations comprising selected receptors to which the adhesins bind.
 XX
 SQ Sequence 900 BP; 208 A; 214 C; 240 G; 238 T; 0 other;
 Alignment Scores:
 Pred. No.: 3 75e-87 Length: 900
 Score: 815.00 Matches: 154
 Percent Similarity: 97.52% Conservative: 3
 Best Local Similarity: 95.65% Mismatches: 4
 Query Match: 96.11% Indels: 0
 DB: 16 Gaps: 0
 US-09-900-575-29_COPY_26_186 (1-161) x AAQ93066 (1-900)
 QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
 DB 139 CCTCCGCGTGAATGTGGGCGCAAAAGCTGTGCTAGATCTTTGACGCAATCTTTGGCAT 198
 QY 21 AsnAspPyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
 DB 199 AACGATTACCCAGAAACCATTAACACTATGTACACTGCACACGAGGTGGCGTTATGGC 258
 QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
 DB 259 GACGTTATCTAGCTTTTCCGGGACCGGTAAATATATGAGCAGTATCTTTCCT 318

```

QY 61 ThrThrsrGIuThrProArGValValTyrAsnSerArgrHsApIysProTtProVal 80
   |||
Db 319 ACTACCGAGGAAAGCGCGGGTGTATTAATTCGAAACGGATTAAGCCGTGGCGGTG 378
QY 81 AlaLeuYrLeuThrProValSerSerAlaGlyLeuValIleYsAlaGlySerLeu 100
   |||
Db 379 GCGCTTATTGACGGCGGTGACAGTGGCGGGAGTGGCGGATTAAGCTGGCTCATTTA 438
QY 101 IleAlaValLeuIleLeuArGInThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
   |||
Db 439 ATTCGCGGTATTATTGCGACACACCAACTATAACGCGATGTATTCACGTTTGTG 498
QY 121 TrpAsnIleYrAlaAsnAsnAspValValAlaProThrGlyGlyCysAspValSerAla 140
   |||
Db 499 TCGAATATTTCAGCCAAATAGATGTGTGTGCTGCCACTGGCGGCTGTGATGTCTGCT 558
QY 141 ArgAspValIleThrValIleThrLeuProAspTyrArgrGlySerValProIleProLeuThrVal 160
   |||
Db 559 CGTGATGTACACGTTACTCTCGGACGACACCTGCTTCAGTGGCGGATTCCTTACCGTT 618
QY 161 Tyr 161
   |||
Db 619 TAT 621

RESULT 40
AA093072 standard; DNA; 900 BP.
ID AA093072 standard; DNA; 900 BP.
XX AA093072;
XX 15-MAR-1996 (first entry)
DE FlmH protein gene from E. coli clinical isolate CI#7.
XX
XX FlmH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW FlmA; FlmF; FlmG; receptor binding site; ss.
XX
XX Escherichia coli clinical isolate CI#7.
XX
FH Key Location/Qualifiers
FT sig_peptide 1..63
FT /*tag= a
FT mat_peptide 64..900
FT /*tag= b
FT
XX W09520657-A1.
PN 03-AUG-1995.
PD 27-JAN-1995; 95WO-DK00042.
XX
XX 27-JAN-1994; 94US-0187166.
PR 27-JAN-1994; 94US-0187166.
XX
XX (GXBI-) GX BIOSYSTEMS AS.
XX
PI Hasly DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX WPI: 1995-275442/36.
DR P-PSDB; AAR76766.
XX
XX Receptor specific bacterial adhesins - useful for targeting active
PT compounds and microbial cells to locations of receptors
XX
XX Example 5; Page 69-74; 152pp; English.
XX
XX The sequences given in AA093062-75 encode FlmH proteins from various E.
CC coli clinical isolates. FlmH is located at the tip of the type 1
CC fimbriae and also intercalated at intervals in the fimbrial organelle.
CC Most forms of the FlmH adhesin target to, and bind to, oligosaccharide
CC structures containing terminally located alpha-D-mannoside residues.
CC FlmH contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison FlmA and
CC the minor components FlmF and FlmG only have 2 cysteine residues.
XX

```

```

CC The localisation of the cysteine residues in FlmH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FlmH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FlmH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of variant FlmH adhesins which
CC may be useful for targeting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.
XX
SQ Sequence 900 BP; 206 A; 215 C; 239 G; 240 T; 0 other;
Alignment Scores:
Pred. No.: 3,75e-87 Length: 900
Score: 815.00 Matches: 155
Percent Similarity: 97.52% Conservative: 2
Best Local Similarity: 96.27% Mismatches: 4
Query Match: 96.11% Indels: 0
DB: 16 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AA093072 (1-900)
QY 1 ProValAlaAsnValGlyGlnAsnLeuValAlaAspLeuSerThrGlnIlePheCysHis 20
   |||
Db 139 CCGCGCGTGAATGGGGCGCAACACCTGTGTGATCTTCGACGCAAACTTTGGCAT 198
QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuIleArgrGlySerAlaTyrGly 40
   |||
Db 199 AACGATTACCCGGAAACCATTAACAGACTATGTCACACTGCACAGGAGTTGGCTTAAGGC 258
QY 41 GlyValLeuSerAsnPhseSerGlyThrValIleYsTyrSerGlySerTyrProPhePro 60
   |||
Db 259 GCGCTGTATCTAATTTTTCGCGACCGCTAAATATGTGGAGTACTTCATTCCT 318
QY 61 ThrThrsrGIuThrProArGValValTyrAsnSerArgrHsApIysProTtProVal 80
   |||
Db 319 ACCACCGAGCAACGCTGCGGTGTATTATTCGAAACGGATTAAGCCGTGGCGGTG 378
QY 81 AlaLeuYrLeuThrProValSerSerAlaGlyLeuValIleYsAlaGlySerLeu 100
   |||
Db 379 GCGCTTATTGACGCGCTGTGACAGTGTGGGGGTGGCGATTAAGCTGGCTCATTTA 438
QY 101 IleAlaValLeuIleLeuArGInThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
   |||
Db 439 ATTCGCGGTATTATTGCGACACACCAACTATAACGCGATGTATTCAGTTTGTG 498
QY 121 TrpAsnIleYrAlaAsnAsnAspValValAlaProThrGlyGlyCysAspValSerAla 140
   |||
Db 499 TCGAATATTTCAGCCAAATAGATGTGTGTGCTGCCACTGGCGGCTGTGATGTCTGCT 558
QY 141 ArgAspValIleThrValIleThrLeuProAspTyrArgrGlySerValProIleProLeuThrVal 160
   |||
Db 559 CGTGATGTACACGTTACTCTCGGACGACACCTGCTTCAGTGGCAATTCCTTACCGTT 618
QY 161 Tyr 161
   |||
Db 619 TAT 621

RESULT 41
AA093065 standard; DNA; 900 BP.
ID AA093065 standard; DNA; 900 BP.
XX AA093065;
XX 15-MAR-1996 (first entry)
DE FlmH protein gene from E. coli clinical isolate CI#4.
XX
XX FlmH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW FlmA; FlmF; FlmG; receptor binding site; ss.
XX

```

```

OS Escherichia coli clinical isolate C1#4.
XX
FH Key Location/Qualifiers
FT sig_peptide 1..63
FT /*tag= a
FT mat_peptide 64..900
FT /*tag= b
XX
XX W09520657-A1.
XX
XX 03-AUG-1995.
XX
XX 27-JAN-1995; 95MO-DK00042.
XX
XX 27-JAN-1994; 94US-0187166.
XX
XX (GXBI-) GX BIOSYSTEMS AS.
XX
XX Haerty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
XX WPI: 1995-275442/36.
XX
XX P-PSDB: AAR76765.
XX
XX Receptor specific bacterial adhesins - useful for targetting active
XX compounds and microbial cells to locations of receptors
XX
XX Example 5; Page 69-74; 152pp; English.
XX
XX The sequences given in AA093062-75 encode FimH proteins from various E.
XX coli clinical isolates. FimH is located at the tip of the type 1
XX fimbriae and also intercalated at intervals in the fimbrial organelle.
XX Most forms of the FimH adhesin target to, and bind to, oligosaccharide
XX structures containing terminally located alpha-D-mannoside residues.
XX FimH contains 4 cysteine residues assumed to direct folding of the
XX molecule into distinct functional domains. For comparison FimA and
XX the minor components FimF and FimG only have 2 cysteine residues.
XX The localisation of the cysteine residues in FimH points to a tandem
XX arrangement of two ancestral genes. Similar amino acids can be
XX found in similar positions in the two halves of the FimH protein. The
XX "midway" point is located roughly around residue 150 in the mature
XX protein. The two halves or domains of FimH have evolved differently
XX with the N-terminal section becoming the domain harbouring the receptor
XX binding site, whereas the C-terminal sector became the domain of the
XX molecule required for integration into the fimbrial organelle. These
XX sequences may be used in the production of variant FimH adhesins which
XX may be useful for targetting active compounds and microbial cells to
XX locations comprising selected receptors to which the adhesins bind.
XX
XX Sequence 900 BP; 205 A; 210 C; 241 G; 244 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 4, 93e-87 Length: 900
XX Score: 814.00 Matches: 154
XX Percent Similarity: 97.52% Conservative: 3
XX Best Local Similarity: 95.65% Mismatches: 4
XX Query Match: 95.99% Indels: 0
XX DB: 16 Gaps: 0
XX
XX US-09-900-575-29_COPY_26_186 (1-161) x AA093065 (1-900)
XX
XX QY 1 ProValValAsnValGlyIAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
XX DB 139 CCTGCGTGAATGCGGCAAAACCTGCGTGCATCTTCGACGCAATCTTTGGCCAT 198
XX QY 21 AsnAspTyrProGlnThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
XX DB 199 AACGATTACCGGAAACCACTTACAGACTATGTCACACTCAGACGAGGTTCCGCTTATGCG 258
XX QY 41 GlyValIleuSerAsnAspSerGlyThrValIleTyrSerGlySerSerTyrProPhePro 60
XX DB 259 GGGGTGTATTCAGATTTTCCGAGACCGTAAATATTAAGGAGTAGTACTATCTTCCCT 318
XX QY 61 ThrThrSerGlnThrProArgValValTyrAsnSerArgThrAspLysProThrProVal 80

```

```

DB 319 ACTACCGCGAAACGCGCGGTGTTATTAATTCGAAACGAGATTAACCGTGGCGGTG 378
QY 81 AlaLeuTyrIleuThrProValSerSerAlaGlyGlyLeuValIleValAlaGlySerIleu 100
DB 379 GCGCTTATTTGACCGCTGTGACAGAGCGGGGAGTGGCGATTAAAGCTGGCTCATTTA 438
QY 101 IleAlaValIleuIleuAsnArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
DB 439 ATGCGCGTCTTATTTTGGCAGACCAACACTATACAGCATGATTTCCAGTTTG 498
QY 121 TrpAsnIleTyrAlaAsnAspValValValProThrGlyGlyAspValSerAla 140
DB 499 TGGATATTTTACCCCAATATATGATGATGATGATGATGATGATGATGATGATGATGAT 558
QY 141 ArgAspValThrValThrIleuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 559 CGTGAATGTCACCGTACTTGTGCGGACTACCGCTGATTCAGTCTCTTACCGTT 618
QY 161 Tyr 161
DB 619 TAT 621
DB
DB 619 TAT 621
DB
RESULT 42
AA093075
ID AA093075 standard; DNA; 900 BP.
XX
XX AA093075;
XX
XX 15-MAR-1996 (first entry)
XX
XX FimH protein gene from E. coli clinical isolate C1#12.
XX
XX FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
XX FimF; FimF; receptor binding site; ss.
XX
XX Escherichia coli clinical isolate C1#12.
XX
XX Key Location/Qualifiers
XX FT sig_peptide 1..63
XX FT /*tag= a
XX FT mat_peptide 64..900
XX FT /*tag= b
XX
XX W09520657-A1.
XX
XX 03-AUG-1995.
XX
XX 27-JAN-1995; 95MO-DK00042.
XX
XX 27-JAN-1994; 94US-0187166.
XX
XX (GXBI-) GX BIOSYSTEMS AS.
XX
XX Haerty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
XX WPI: 1995-275442/36.
XX
XX P-PSDB: AAR76764.
XX
XX Receptor specific bacterial adhesins - useful for targetting active
XX compounds and microbial cells to locations of receptors
XX
XX Example 5; Page 69-74; 152pp; English.
XX
XX The sequences given in AA093062-75 encode FimH proteins from various E.
XX coli clinical isolates. FimH is located at the tip of the type 1
XX fimbriae and also intercalated at intervals in the fimbrial organelle.
XX Most forms of the FimH adhesin target to, and bind to, oligosaccharide
XX structures containing terminally located alpha-D-mannoside residues.
XX FimH contains 4 cysteine residues assumed to direct folding of the
XX molecule into distinct functional domains. For comparison FimA and
XX the minor components FimF and FimG only have 2 cysteine residues.
XX The localisation of the cysteine residues in FimH points to a tandem

```

OS	Escherichia coli B212
XX	
KW	therapy; vaccine; vir

Db 316 GCGCTTATTGACGCCCTGTGAGCAGTGGGGGAGTGGCGATTAAAGCAGGCTCATTA 37

Db 316 GCGCTTATTGACGCCCTGTGAGCAGTGGGGGAGTGGCGATTAAAGCAGGCTCATTA 37

QY 101 ILEAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
 Db 376 ATGGCGGTGCTATTTGGACAGACCAACATATACAGCATGATTTCCAGTTTG 435
 QY 121 TTPAsnIleTyrAlaAsnAsnAspValValProthGlyGlyCysAspValSeraLa 140
 Db 436 TGGAAATATTACGCCCAATATGATGTGTGTCGCCACGTGGCGGCTGCATCTCTGCT 495
 QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
 Db 496 CGTGATGTCACCGTCTACTGCGGACATCCTGCTGTCAGTCCGATTCCTTACCGTT 555
 QY 161 Tyr 161
 Db 556 TAT 558

RESULT 44

AAD29370
 ID AAD29370 standard: DNA; 837 BP.

AC AAD29370;

DT 07-MAY-2002 (first entry)

DE Escherichia coli strain EC95 FimH DNA.

KM FimH: Immune response; antibacterial; enterobacillus-related disease;
 therapy: vaccine; urinary tract infection; bladder; ds.

OS Escherichia coli EC95.

Key Location/Qualifiers
 CDS 1..837

FT /*tag- a
 FT /product- "E. coli FimH protein"
 FT /transl_except- (pos:526..528, aa:Arg)
 FT /transl_except- (pos:601..603, aa:His)
 FT /note- "CDS does not include start and stop codon"
 FT /partial

PN MO200204496-A2.

PD 17-JAN-2002.

PF 06-JUL-2001: 2001MO-US21525.

PR 07-JUL-2000: 2000US-216750P.

PA (MEDI-) MEDIMMUNE INC.

PI Langermann S, Revel A, Auguste C, Burieln J;

DR WPI: 2002-111702/22.

DR P-PSDB: AAE18431.

PT New immunogenic polypeptide, useful as vaccine for protecting against
 an enterobacillus-related disease in a patient at risk of contracting
 such disease, e.g. urinary tract infection or a bladder infection
 Example 1: Fig 1; 101pp; English.

CC The invention relates to bacterial immunogenic agents for administration
 CC to humans and non-human animals to stimulate an immune response. The
 CC invention also relates to methods for vaccination of mammalian species
 CC with variants of E. coli FimH protein derived from different strains of
 CC E. coli. The vaccine composition or the antibody is useful for protecting
 CC against and treating an enterobacillus-related disease in a patient
 CC afflicted or at a risk of contracting the disease. In particular, the
 CC disease is a urinary tract or bladder infection. The disease is caused
 CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
 CC The present sequence is Escherichia coli strain EC95 FimH DNA.

Sequence 837 BP; 194 A; 199 C; 223 G; 221 T; 0 other;
 Alignment Scores:
 Pred. No.: 7,7e-87
 Score: 812.00
 Percent Similarity: 97.52%
 Best local Similarity: 95.65%
 Query Match: 95.75%
 Ds: 24
 Gaps: 0
 US-09-900-575-29_COPY_26_186 (1-161) x AAD29370 (1-837)

QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
 Db 76 CTTCCGCTGATGTGGGCAAAACCTGCTGATGATCTTGCAGCGCAAACTTTGGCATT 135
 QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuIleArgGlySeraLaTyrGly 40
 Db 136 AAGCATACCAGCAAAACCATTTACGACTATGTACACACTGCACAGAGGTTCCGCTTATGCC 195
 QY 41 GlyValLeuSerAsnAspPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
 Db 196 GCGCTGTATCTAGTATTTTCCGGACCGTAATAATATAGCGCATATCTTCCCT 255
 QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProThrProVal 80
 Db 256 ACTACACGCGAAACGCGCGGTTGTTAAATTCAGAAAGGATTAAGCGCGCGGTG 315
 QY 81 AlaLeuTyrLeuThrProValSerSeraLaGlyLeuValIleLysAlaGlySeraLeu 100
 Db 316 GCGCTTATTTGACGCTGTGACACTGCGGGGAGTGCCGATTAACCTGCTCATTA 375
 QY 101 ILEAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
 Db 376 ATGGCGGTGCTATTTGGACAGACCAACATATACAGCATGATTTCCAGTTTG 435
 QY 121 TTPAsnIleTyrAlaAsnAsnAspValValProthGlyGlyCysAspValSeraLa 140
 Db 436 TGGAAATATTACGCCCAATATGATGTGTGTCGCCACGTGGCGGCTGCATCTCTGCT 495
 QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
 Db 496 CGTGATGTCACCGTCTACTGCGGACATCCTGCTGTCAGTCCGATTCCTTACCGTT 555
 QY 161 Tyr 161
 Db 556 TAT 558

RESULT 45

AAQ93069
 ID AAQ93069 standard.

AC AAQ93069;

DT 15-MAR-1996 (first)

DE FimH protein gene

KW FimH: type 1 fimbria

KW FimH: FimH; FimH;

OS Escherichia coli clonurear isolate C1#10.

Key Location/Qualifiers
 CDS 1..63

FT mat_peptide

FT /*tag- a
 FT /*tag- b

PN MO9520657-A1.

PD 03-AUG-1995.

Added to file
confirmed
USP

0.
 pmoside residue;

PF 27-JAN-1995: 95MO-DK0042.
XX
PR 27-JAN-1994: 9AUS-0187166.
XX
XX (GXBI-) GX BIOSYSTEMS AS.
XX
PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
DR WPI, 1995-275442/36.
DR P-PSDB; AAR/6767.

Receptor specific bacterial adhesins - useful for targeting active compounds and microbial cells to locations of receptors

Example 5; Page 69-74; 152pp; English.

The sequences given in AA093062-75 encode FimH proteins from various E. coli clinical isolates. FimH is located at the tip of the type 1 fimbriae and also intercalated at intervals in the fimbrial organelle. Most forms of the FimH adhesin target to, and bind to, oligosaccharide structures containing terminally located alpha-D-mannoside residues. FimH contains 4 cysteine residues assumed to direct folding of the molecule into distinct functional domains. For comparison FimA and the minor components FimF and FimG only have 2 cysteine residues. The localisation of the cysteine residues in FimH points to a tandem arrangement of two ancestral genes. Similar amino acids can be found in similar positions in the two halves of the FimH protein. The "midway" point is located roughly around residue 150 in the mature protein. The two halves or domains of FimH have evolved differently with the N-terminal section becoming the domain harbouring the receptor binding site, whereas the C-terminal sector became the domain of the molecule required for integration into the fimbrial organelle. These sequences may be used in the production of variant FimH adhesins which may be useful for targeting active compounds and microbial cells to locations comprising selected receptors to which the adhesins bind.

Sequence 888 BP; 203 A; 211 C; 234 G; 240 T; 0 other;

Alignment Scores:

Pred. No.: 3,296-86 Length: 888
Score: 807.00 Matches: 155
Percent Similarity: 96.27% Conservative: 0
Best Local Similarity: 96.27% Mismatches: 2
Query Match: 95.17% Indels: 4
DB: 16 Gaps: 1

US-09-900-575-29_copy_26_186 (1-161) x AA093069 (1-888)

QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 139 CCGCGCGTGAATGTGGGCAAACTGGTCGTGATCTTTCGACGCAAACTTTGGCCAT 198
QY 21 AsnAspTyrProGlnThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 199 AACGATTATCCGGAACCATTAACACTATGTCACACTGCAACGAGCTCGGCTTATGCG 258
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
DB 259 GCGGTGTATCTAATTTTCCGGACCGTAAATATATAGTGGCAGTACATTCATTCCCT 318
QY 61 ThrThrSerGlnThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
DB 319 ACCACCGACGAAACCCGCGCGTGTATTATTCGAGAAAGGATTAAGCCGTGCGCGTG 378
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
DB 379 GCGTATTATTGACGCGCTGTGACAGTGGGT-----AAAGCTGGCTCATTA 426
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnTyrAsnSerAspAspPheGlnPheVal 120
DB 427 ATTGCGGTGTATTGTGCGACACCACTATACAGCGATGATTTCCAGTTTG 486
QY 121 TrpAsnIleTyrAlaAsnAspValValValProThrGlyGlyCysAspValSerAla 140

DB 487 TGGAAATTATACGCCAATATATATGATGTGGTGGCTACTGGGGGTGCGATGTTCTGCT 546
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 547 CCGGATGTCACCGTTACTCTGCGGACTACCTGGTTCAGTCCATTCCTTACCGTT 606
QY 161 Tyr 161
DB 607 TAT 609

Search completed: November 28, 2002, 19:14:35
Job time : 267 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 28, 2002, 19:06:46 ; Search time 2576 Seconds

(without alignments)
1818.925 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186

Perfect score: 848

Sequence: 1 PVTNNGQNLVVDLSQIFCH.....DVTWLPDRCGVPPIPLTV 161

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n_model -DEV=x1h
-O=/cgn2.1/USPTO_spool/US09900575/runat.22112002.130708.4545/app-query.fasta.1.327
-DB=GenEmbl -GEMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=Blasum62 -TRANS=human40.cdi -LIST=45
-OUTMT=ptc -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=45 -MODE=LOCAL
-USER=US09900575_eegN.1.1-1516.0runat.22112002.130708.4545 -NCPU=6 -ICPU=3
-NO_XLPRX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARM.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sbs: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sbs: *
28: em_un: *

29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pla: *
35: em_htg_rod: *
36: em_htg_mem: *
37: em_htg_vtl: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	837	6	AX354041
2	841	99.2	837	6	AX354051
3	841	99.2	2050	1	ECRIMEGH
4	834	98.3	837	6	AX354040
5	834	98.3	837	6	AX354055
6	834	98.3	840	6	AX354044
7	834	98.3	840	6	AX354046
8	834	98.3	903	6	AX354082
9	834	98.3	903	6	AX363712
10	834	98.3	1040	1	AF154925
11	834	98.3	1040	1	AF154925
12	834	98.3	4637	6	AX357443
13	834	98.3	7416	6	AX354080
14	834	98.3	9299	6	AX357444
15	834	98.3	11313	1	ECODW93
16	834	98.3	338534	6	AX354037
17	831	98.0	837	6	AX354037
18	830	97.9	837	6	AX354037
19	828	97.5	1040	1	AF154928
20	827	97.5	1040	1	AF306535
21	826	97.4	837	6	AX354049
22	826	97.4	1040	1	AF306536
23	825	97.3	840	6	AX354039
24	824	97.2	837	6	AX354054
25	824	97.2	14466	1	AE005662
26	824	97.2	168764	1	AP002569
27	823	97.1	1040	1	AF154927
28	822	96.9	837	6	AX354050
29	822	96.9	837	6	AX354052
30	822	96.9	937	1	AF288194
31	820	96.7	837	6	AX354042
32	819	96.6	837	6	AX354045
33	819	96.6	837	6	AX354056
34	819	96.6	840	6	AX354038
35	819	96.6	840	6	AX354038
36	819	96.6	840	6	AX354038
37	819	96.6	1008	1	AF089840
38	818	96.5	1040	1	AF306537
39	817	96.3	837	6	AX354035
40	816	96.2	837	6	AX354047
41	816	96.2	837	6	AX354048
42	816	96.2	837	6	AX354048
43	812	95.8	837	6	ECRTM
44	812	95.8	837	6	AX354036
45	807	95.2	1039	1	AF154929

RESULT 1

ALIGNMENTS

AX354041
LOCUS AX354041 837 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 7 from Patent WO0204496.
ACCESSION AX354041
VERSION AX354041.1 GI:18618978
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE
1 Langermann, S., Revel, A., Auguste, C. and Burllein, J.
Fimb adhesin proteins and methods of use
Patent: WO 0204496-A 7 17-JAN-2002;
JOURNAL MEDIMUNE, INC. (US)
MEDIMUNE, INC. (US)
Location/Qualifiers
1. 837
/organism="Escherichia coli"
/db_xref="taxon:562"

BASE COUNT 193 a 200 c 220 g 224 t
ORIGIN

Alignment Scores:
Pred. No.: 3.4e-78 Length: 837
Score: 848.00 Matches: 161
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-09-900-575-29_copy_26_186 (1-161) x AX354041 (1-837)

QY 1 ProvalValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 76 CCCGTCGTGAATGTGGGGCAAAACCTGTCGATCTTTCAGCCAAATCTTTGGCAT 135

QY 21 AsnAspTyProGluThrIleThrAspTyValThrLeuGlnArgIleSerAlaTyrgly 40
DB 136 AACGATTATCCGGAACCAATTCAGACTATGTCACACTGCAACAGAGCTCGGCTTATGCG 195

QY 41 GlyValLeuSerAsnPheserGlyThrValLysTySerGlySerSerTyProPhePro 60
DB 196 GGGCTGTATCTAATTTTTCGCGACCGTAATAATAGTGGCAGTATCCATTCTCT 255

QY 61 ThrThrSerGluThrProArgValValTyAsnSerArgThrAspLysProThrProVal 80
DB 256 ACCACGACGAAACGCCGCGCTGTTTAAATTCAGAAACGATTAAGCCGTGGCGGTG 315

QY 81 AlaLeuTyrlleuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
DB 316 GCGCTTATTTGACGCTGTGACAGTGGCGGGTGGTGTATTAAGCTGGCTCATTA 375

QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyAsnSerAspPheGlnPheVal 120
DB 376 ATTCGCGCTTATTTTTCGACAGACCAACACTATAACGATATTCAGTTGTG 435

QY 121 TrpAsnIleTyAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
DB 436 TGGATATATTAGCCCAATAATGATGTGTGTGCTACTGCGCGCTCCGATGTTTGGCT 495

QY 141 ArgAspValThrValThrLeuProAspTyArgGlySerValProIleProLeuThrVal 160
DB 496 CGTGTATGCACGTTACTCTGCGGACTACCGTGTTCAGTGCACAAATTCCTTACCGTT 555

QY 161 Tyr 161
DB 556 TAT 558

RESULT 2
LOCUS AX354051 837 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 17 from Patent WO0204496.
ACCESSION AX354051

VERSION AX354051.1 GI:18618988
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE
1 Langermann, S., Revel, A., Auguste, C. and Burllein, J.
Fimb adhesin proteins and methods of use
Patent: WO 0204496-A 17 17-JAN-2002;
JOURNAL MEDIMUNE, INC. (US)
MEDIMUNE, INC. (US)
Location/Qualifiers
1. 837
/organism="Escherichia coli"
/db_xref="taxon:562"

BASE COUNT 191 a 200 c 223 g 223 t
ORIGIN

Alignment Scores:
Pred. No.: 1.79e-77 Length: 837
Score: 841.00 Matches: 159
Percent Similarity: 99.38% Conservative: 1
Best Local Similarity: 98.76% Mismatches: 1
Query Match: 99.17% Indels: 0
Gaps: 0

US-09-900-575-29_copy_26_186 (1-161) x AX354051 (1-837)

QY 1 ProvalValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 76 CCCGTCGTGAATGTGGGGCAAAACCTGTCGATCTTTCAGCCAAATCTTTGGCAT 135

QY 21 AsnAspTyProGluThrIleThrAspTyValThrLeuGlnArgIleSerAlaTyrgly 40
DB 136 AACGATTATCCGGAACCAATTCAGACTATGTCACACTGCAACAGAGCTCGGCTTATGCG 195

QY 41 GlyValLeuSerAsnPheserGlyThrValLysTySerGlySerSerTyProPhePro 60
DB 196 GGGCTGTATCTAATTTTTCGCGACCGTAATAATAGTGGCAGTATCCATTCTCT 255

QY 61 ThrThrSerGluThrProArgValValTyAsnSerArgThrAspLysProThrProVal 80
DB 256 ACCACGACGAAACGCCGCGCTGTTTAAATTCAGAAACGATTAAGCCGTGGCGGTG 315

QY 81 AlaLeuTyrlleuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
DB 316 GCGCTTATTTGACGCTGTGACAGTGGCGGGTGGTGTATTAAGCTGGCTCATTA 375

QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyAsnSerAspPheGlnPheVal 120
DB 376 ATTCGCGCTTATTTTTCGACAGACCAACACTATAACGATATTCAGTTGTG 435

QY 121 TrpAsnIleTyAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
DB 436 TGGATATATTAGCCCAATAATGATGTGTGTGCTACTGCGCGCTCCGATGTTTGGCT 495

QY 141 ArgAspValThrValThrLeuProAspTyArgGlySerValProIleProLeuThrVal 160
DB 496 CGTGTATGCACGTTACTCTGCGGACTACCGTGTTCAGTGCACAAATTCCTTACCGTT 555

QY 161 Tyr 161
DB 556 TAT 558

RESULT 3
LOCUS ECFIMFGH 2050 bp DNA linear BCT 12-SEP-1993
DEFINITION E.coli genes fimb, fimg and fimH.
ACCESSION X05672
VERSION X05672.1 GI:41463
KEYWORDS fimb gene; fimg gene; fimH gene; regulatory protein.
SOURCE Escherichia coli.
ORGANISM Escherichia coli

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 2050)
 Klemm P. and Christiansen G.
 Three fim genes required for the regulation of length and mediation
 of adhesion of Escherichia coli type 1 fimbriae
 JOURNAL
 Mol. Gen. Genet. 208 (3), 439-445 (1987)
 MEDLINE
 88038337
 PUBMED
 2890081
 FEATURES
 source
 location/Qualifiers
 1. 2050
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 17. 21
 /note="ribosome binding site (f1mF)"
 28. 558
 /note="pot. f1mF precursor (AA 1-176)"
 /codon_start=1
 /transl_table=1
 /protein_id="CAA29154.1"
 /db_xref="GI:41464"
 /db_xref="SWISS-PROT:P08189"
 /translation="MKRKPYLLCAFLMLAVSHALADSTTTTRGYVRDNGSCVAAS
 TNFTVDLMENAKOFNNIGATTPVPSFRILLPCGVNVSRRKGMVYWRADSHANLILA
 LENTASASGLIGLQLENEQNOIPLNAPSALLSWTLTPGKPTMLNFTARLMATQVPIV
 TAGHINATATFTLEYQ"
 87. 88
 /note="pot. processing site of the mature protein"
 561. 565
 /note="ribosome binding site (f1mG)"
 571. 1074
 /note="pot. f1mG precursor (AA 1-167)"
 /codon_start=1
 /transl_table=1
 /protein_id="CAA29155.1"
 /db_xref="GI:41465"
 /db_xref="SWISS-PROT:P08190"
 /translation="MKCKRGVLAAILALASATTGAADVTITNGKVVAKPCTVSTT
 NATVDLGDIVSFLMSGASAWHDVALDELTCNPGVTSRYTASFSGADSTGYKNG
 TAONIQLLELDDSGNLTNGATKTVQVDDSSQSAHPPLQVRAIVLVGANTQGIQAVI
 SIITYYS"
 639. 640
 /note="pot. processing site of the mature protein"
 1081. 1085
 /note="ribosome binding site (f1mH)"
 1094. 1996
 /note="pot. f1mH precursor (AA 1-300)"
 /codon_start=1
 /transl_table=1
 /protein_id="CAA29156.1"
 /db_xref="GI:41466"
 /db_xref="SWISS-PROT:P08191"
 /translation="MKRVITLFAVLMLGSVNAVSPACKTANGTAIPIGGSANYVN
 LAHVNVGONLVLDISTOJCHNDYPERITDVTLOVRSAGVLSNPGYKYSGSS
 YPEPTSETPRVYNSTRDKPMKVALXYLTPVSAGVAKKASGLAVALLKRTNNYS
 DDQRFYMNITYKANDDYVPTGCGDVSAKRDVTLVLPDRGSVPILVLYAKRSQNLGYL
 SGRHADGASITNTASFSPACGVGLTRNGTIIIPANNVSLGAVGSAVSLGLTAN
 YARTGGQVYAGNVSLIGVTVYQ"
 1162. 1163
 /note="pot. processing site of the mature protein"

BASE COUNT 494 a 489 c 542 g 525 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5, 01e-77 Length: 2050
 Score: 841.00 Matches: 159
 Percent Similarity: 99.38% Conservative: 1
 Best Local Similarity: 98.76% Mismatches: 1
 Query Match: 99.17% Indels: 0
 DB: 1 Gaps: 0
 US-09-900-575-29_copy_26_186 (1-161) x ECFIMFGH (1-2050)

Qy 1 ProValValasnaValGlyGlnasnlLeuValValaspleuserThrgnIlePheCySHs 20
 Db 1232 CCCGTCGATGATGGGGCAAAACCTGCTGTCGATCTTCGACGCAAACTTTTGCAT 1291
 Qy 21 AsnspYrProGluThrIleThraSPYrValThrLeuGlnargGlySeraIaTyrgly 40
 Db 1292 AACGATTAATCCGAAACCATTAACGACTATGTACACTCCACGAGAGCTCGCTATAGC 1351
 Qy 41 GlyValLeuserAsnPheserGlyThrValIlystYserGlyserSerTyrrProPhePro 60
 Db 1352 GCGCTTATCTATCTATTTTCGCGGACCGTAAATATAGTGGAGTACCTATTCATTCCT 1411
 Qy 61 ThrThrserGluThrProArGvalValTyraSnsArGThraspYsProTrProVal 80
 Db 1412 AACACCGACGAAACCGCCGCGTGTATATATTCGAGAACGATTAACCGCGCGG 1471
 Qy 81 AlaLeuTyrrLeuThrProValSerSerAlaGlyGlyLeuValIleIysAlaGlySerLeu 100
 Db 1472 GCGCTTATTTGACGCTGTAGCAGAGTGGCGGGGCGGATTAACCTGCTCATTA 1531
 Qy 101 IleAlaValLeuIleLeuArGlnThraSnsArGThraspYsSerAspPheGlnPheVal 120
 Db 1532 ATTAGCGGTGCTTATTTTGGCAGACCAACTATTAACGCGATATTTCCAGTTTGTG 1591
 Qy 121 TrpAsnIleTyrrAlaasnsAspValValProThrGlyGlyCysAspValSeraIa 140
 Db 1592 TGGAAATATTTACGCCAATATGATGTGTGTCCTAACGCGGCGGATGTTCTGCT 1651
 Qy 141 ArgAspValThrValThrLeuProAspTyrrArgGlySerValProIleProLeuThrVal 160
 Db 1652 CGTGATGTCAACCGTTACTGTCCGCGACTACGHTGTTCACTGCATTCCTTACCGTT 1711
 Qy 161 Tyr 161
 Db 1712 TAT 1714
 RESULT 4
 LOCUS AX354040 837 bp DNA linear PAT 06-FEB-2002
 DEFINITION Sequence 6 from Patent WO0204496.
 ACCESSION AX354040
 VERSION AX354040.1 GI:18618977
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli.
 Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1
 Langemann, S., Revel, A., Auguste, C. and Burelein, J.
 Fimbriin adhesin proteins and methods of use
 Patent: WO 0204496-A 6 17-JAN-2002;
 MEDLINE
 JOURNAL
 FEATURES
 source
 location/Qualifiers
 1. 837
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 BASE COUNT 192 a 202 c 223 g 220 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 9.4e-77 Length: 837
 Score: 834.00 Matches: 158
 Percent Similarity: 98.76% Conservative: 1
 Best Local Similarity: 98.14% Mismatches: 2
 Query Match: 98.35% Indels: 0
 DB: 6 Gaps: 0
 US-09-900-575-29_copy_26_186 (1-161) x AX354040 (1-837)
 Qy 1 ProValValasnaValGlyGlnasnlLeuValValaspleuserThrgnIlePheCySHs 20
 Db 76 CCCGCGGATGATGGGGCAAAACCTGCTGTCGATCTTCGACGCAAACTTTTGCAT 135

QY	136	AAAGATTATCCGGAACCACTTACAGACTATGTCACACGACGAGGCTCGGCTTATGGC	195
QY	21	AsnAspTyrProGluThrIleThrAspTyrValIThrLeuGlnArgGlySerAlaTyrGly	40
Db	136	AAAGATTATCCGGAACCACTTACAGATTATGTCACACGACGAGGCTCGGCTTATGGC	195
QY	41	GlyValLeuSerAsnHisSerGlyThrValIlysTyrSerGlySerSerTyrProPhePro	60
Db	196	GGGCTGTATTACTAATTTTTCGGGACCGTAAATATAGGCGAGTAAAGCTATTCATTTCCG	255
QY	61	ThrTyrSerGluThrProAlaArgValIlyrAsnSerArgThrAspLysProTyrProVal	80
Db	256	ACCACCGATGAACCGCGCGGGTGTATTATATTCGAGAACGATTAACCGCGGCGGTG	315
QY	81	AlaLeuTyrIleuThrProValIserSerAlaGlyIleuValIleAlaValaGlySerLeu	100
Db	316	GGCGTTTATTTGACCGCTGTGACCAATGGCGGGGGGTGTATTAAAGCTGGCTCATTA	375
QY	101	IleAlaValLeuIleLeuArgIlnThrAsnAspTyrAsnSerSerAspPheGlnPheVal	120
Db	376	ATTCCCTGCTTATTTTGGCAGACCAACACTATTAACAGGATGATTTCCAGTTTGG	435
QY	121	TrpAsnIleTyrAlaAsnAspValValProThrGlyGlyCysAspValSerAla	140
Db	436	TGGAATATTTTACGCCAATATGATGTGGGTGCCACACGTGGCGCTGATTTCTGCT	495
QY	141	ArgAspValThrValIThrLeuProAspTyrArgGlySerValProIleProLeuThrVal	160
Db	496	CGTATGTACACCGTTACTCTGCCGACTACCTCGTTGATGCGGCAATCTCTTACCGTT	555
QY	161	Tyr	161
Db	556	TAT	558
RESULT 5			
LOCUS	AX354055	837 bp	DNA linear PAT 06-FEB-2002
DEFINITION	Sequence 21 from Patent WO0204496.		
ACCESSION	AX354055		
VERSION	AX354055.1	GI:18618992	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	1		
TITLE	Langermann, S., Revel, A., Auguste, C., and Burlein, J.		
JOURNAL	Film adhesin proteins and methods of use		
FEATURES	Patent: WO 0204496-A 21 17-JAN-2002;		
source	MEDIMUNE, INC. (US)		
	Location/Qualifiers		
	1..837		
	/organism="Escherichia coli"		
	/db_xref="taxon:562"		
BASE COUNT	191 a 201 c 222 g 223 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	9.4e-77	Length:	837
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	6	Gaps:	0
US-09-900-575-29_COPY_26_186 (1-161) x AX354055 (1-837)			
QY	1	ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis	20
Db	76	CCCGTCGTGAAGTGGGGGAAAACCTGCTCGATCTTTTCGACGCAAACTTTTGGCAT	135
QY	21	AsnAspTyrProGluThrIleThrAspTyrValIThrLeuGlnArgGlySerAlaTyrGly	40
Db	136	AAAGATTATCCGGAACCACTTACAGACTATGTCACACGACGAGGCTCGGCTTATGGC	195

QY	41	glyValleuserAsnpheserGlyThrvAllysyrserserTyrrProphPro	60
Db	196	GACCGTATCTAATTATTTTCGGGAGCCCTAAATAATAGTGCATAGCTATCCATTTCCT	255
QY	61	ThrTrSerGluThrProArgValValTyraSenserarqThraSplysProTrpProVal	80
Db	256	ACCACACGAAACCGCCGCGCTTCTTATATTCGAAACGATPAAGCCGTGCGCGTG	315
QY	81	AlaLeuTyrrLeuThrProValSerSerAlaGlyLeuValIleIysAlaGlySerLeu	100
Db	316	GCGCTTATTATGACCCCTGTGAGCAGTACGGCGCGCGGCGGATTAAGCGGCTCATTA	375
QY	101	IleAlaValLeuIleLeuArgGlnThrAsnSntTyraSenserAspPheGlnPheVal	120
Db	376	ATTGGCGGCTTATTTTCGAGACGACCAACATTAACAGCATGATTTCCAGTTTGTG	435
QY	121	TrpAsnIleTyrrAlaAsnAsnAspValValProThrgIglyCysAspValSerAla	140
Db	436	TGGATATTATTCGCCAATATATGATGTGGTGGCTACTGGCGGCTCGAGTGTTCCT	495
QY	141	ArgAspValThrValThrLeuProAspTyrrArgGlySerValProIleProLeuThrVal	160
Db	496	CCTGATGTCACCGTTACTCTGCCGAGCAACCTCGGTTCAAGTGCATTCCTTAACGCT	555
QY	161	Tyr 161	
Db	556	TAT 558	
RESULT 6	AX354044	840 bp	DNA linear PAT 06-FEB-2002
LOCUS	AX354044	Sequence 10 from Patent WO0204496.	
DEFINITION	AX354044		
ACCESSION	AX354044.1	GI:18618981	
VERSION	AX354044.1	GI:18618981	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE	1		
AUTHORS	Langermann, S., Revel, A., Auguste, C. and Butlein, J.		
TITLE	Film adhesin proteins and methods of use		
JOURNAL	Patent: WO 0204496-A 10 17-JAN-2002;		
FEATURES			
source	1..840	location/Qualifiers	
BASE COUNT	193 a 202 c 224 g 221 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	9, 44e-77	length:	840
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	6	Gaps:	0
US-09-900-575-29_COPY_26_186 (1-161) x AX354044 (1-840)			
QY	1	proValValAsnValGlyIlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis	20
Db	76	CCCGCGGTGATGATGGGGGCAAAACCTGTGCTGTGATCTTCGACGCAAAATCTTTTCCAT	135
QY	21	AsnAspTyrrProGluThrIleThrAspTyrrValThrLeuIlnArgGlySerAlaTyrrGly	40
Db	136	AACCATTAACCCGGAACCATTAACGATTATGTACACATGCAACGAGGCTGCGCTTATGCG	195
QY	41	glyValleuserAsnpheserGlyThrvAllysyrserserTyrrProphPro	60
Db	196	GACCGTATCTAATTATTTTCGGGAGCCCTAAATAATAGTGCATAGCTATCCATTTCCT	255

QY 61 ThThSergluthrProArqVal1TyraSnsSerArgThraSplysProtrProval 80
Db 256 ACCACAGTGAAGCGCGGGTGTATTAATTCGAGAACGATACCGCTGGCCGGTG 315
QY 81 AlaleuTyrluThrProvalSerSeraIaglyLeuValIlelysaIaglySerleu 100
Db 316 GCGCTTATTGACGCCCTGTGACAGTGGCGGGGGGATTAAGCTGGCTCATTTA 375
QY 101 IleAlaValleuIleleuArqInThraSnsSerArgThraSplysProtrProval 120
Db 376 ATGGCGCTGCTTATTGGACAGACCAACTATACAGCGATGATTCAGTTTGTG 435
QY 121 TrpaSnIleTyraIlaSnsAspValVal1ProthrglyGlyCysAspValSeraIa 140
Db 436 TGGATATTACCGCAATATGATGTGGTGGCCAGTGGCGGCTGGATTTCTGCT 495
QY 141 ArgAspValThrValThreuproAspyrArgGlySerValProIleProleuThrVal 160
Db 496 CGTGATTCACCGCTTACTGTGCGGACTACCGTGTTCAGTTCCTCTTACCGTT 555
QY 161 Tyr 161
Db 556 TAT 558
RESULT 7
AX354046 840 bp DNA linear PAT 06-FEB-2002
LOCUS AX354046
DEFINITION Sequence 12 from Patent WO0204496.
ACCESSION AX354046
VERSION AX354046.1 GI:18618983
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1
AUTHORS Langermann, S., Revel, A., Auguste, C. and Burieln, J.
TITLE Film adhesin proteins and methods of use
JOURNAL Patent: WO 0204496-A 12 17-JAN-2002;
MEDIMUNE, INC. (US)
FEATURES
source Location/Qualifiers
1..840
/organism="Escherichia coli"
/db_xref="taxon:562"
BASE COUNT 193 a 201 c 222 g 224 t
ORIGIN
Alignment Scores:
Pred. No.: 9.44e-77 Length: 840
Score: 834.00 Matches: 158
Percent Similarity: 98.76% Conservative: 1
Best Local Similarity: 98.14% Mismatches: 2
Query Match: 98.35% Indels: 0
Gaps: 0
US-09-900-575-29_COPY_26_186 (1-161) x AX354046 (1-840)
QY 1 ProvalValaSnValGlyGlnaSnleuValaAspleuSerThrgInIlePhcysHls 20
Db 76 CCGCTGCTGAATGTGGGCAAAACCTGTGCTGATCTTTCAGCCCAATCTTTGCCAT 135
QY 21 AsnaSptyrProgluThrlleThraSptyrValThrlengInaArglySeraIaTyrgly 40
Db 136 AACGATTATCCGGAACCATTAACAGACTATGTCACACTGCAAGAGGCTCGGCTTATGCG 195
QY 41 GlyValleuSerAsnPhseSerglyThrValIlystyrsSerglySeraSertyrProphero 60
Db 196 GCGCTTATTGATTTTCGCGGACCGTAATAATATATAGTGGCAGTATCATTTCT 255
QY 61 ThThSergluthrProArqVal1TyraSnsSerArgThraSplysProtrProval 80
Db 256 ACCACAGTGAAGCGCGGGTGTATTAATTCGAGAACGATACCGCTGGCCGGTG 315

QY 81 AlaleuTyrluThrProvalSerSeraIaglyLeuValIlelysaIaglySerleu 100
Db 316 GCGCTTATTGACGCCCTGTGACAGTGGCGGGGGGATTAAGCTGGCTCATTTA 375
QY 101 IleAlaValleuIleleuArqInThraSnsSerArgThraSplysProtrProval 120
Db 376 ATGGCGCTGCTTATTGGACAGACCAACTATACAGCGATGATTCAGTTTGTG 435
QY 121 TrpaSnIleTyraIlaSnsAspValVal1ProthrglyGlyCysAspValSeraIa 140
Db 436 TGGATATTACCGCAATATGATGTGGTGGCCAGTGGCGGCTGGATTTCTGCT 495
QY 141 ArgAspValThrValThreuproAspyrArgGlySerValProIleProleuThrVal 160
Db 496 CGTGATTCACCGCTTACTGTGCGGACTACCGTGTTCAGTTCCTCTTACCGTT 555
QY 161 Tyr 161
Db 556 TAT 558
RESULT 8
AX354082 903 bp DNA linear PAT 06-FEB-2002
LOCUS AX354082
DEFINITION Sequence 48 from Patent WO0204496.
ACCESSION AX354082
VERSION AX354082.1 GI:18618996
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1
AUTHORS Langermann, S., Revel, A., Auguste, C. and Burieln, J.
TITLE Film adhesin proteins and methods of use
JOURNAL Patent: WO 0204496-A 48 17-JAN-2002;
MEDIMUNE, INC. (US)
FEATURES
source Location/Qualifiers
1..903
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Sequence 396 film plus native signal sequence"
BASE COUNT 206 a 213 c 240 g 244 t
ORIGIN
Alignment Scores:
Pred. No.: 1.03e-76 Length: 903
Score: 834.00 Matches: 158
Percent Similarity: 98.76% Conservative: 1
Best Local Similarity: 98.14% Mismatches: 2
Query Match: 98.35% Indels: 0
Gaps: 0
US-09-900-575-29_COPY_26_186 (1-161) x AX354082 (1-903)
QY 1 ProvalValaSnValGlyGlnaSnleuValaAspleuSerThrgInIlePhcysHls 20
Db 139 CCGCTGCTGAATGTGGGCAAAACCTGTGCTGATCTTTCAGCCCAATCTTTGCCAT 198
QY 21 AsnaSptyrProgluThrlleThraSptyrValThrlengInaArglySeraIaTyrgly 40
Db 199 AACGATTATCCGGAACCATTAACAGACTATGTCACACTGCAAGAGGCTCGGCTTATGCG 258
QY 41 GlyValleuSerAsnPhseSerglyThrValIlystyrsSerglySeraSertyrProphero 60
Db 259 GCGCTTATTGATTTTCGCGGACCGTAATAATATATAGTGGCAGTATCATTTCT 318
QY 61 ThThSergluthrProArqVal1TyraSnsSerArgThraSplysProtrProval 80
Db 319 ACCACAGTGAAGCGCGGGTGTATTAATTCGAGAACGATACCGCTGGCCGGTG 378
QY 81 AlaleuTyrluThrProvalSerSeraIaglyLeuValIlelysaIaglySerleu 100
Db 379 GCGCTTATTGACGCCCTGTGACAGTGGCGGGGATTAAGCTGGCTCATTTA 438

Qy	101	IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal	120
Db	439	ATTGGCGGCTTATTGTCGACAGACCACAACTATAACACGATGATTTCCAGTTGTG	498
Qy	121	TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla	140
Db	499	TGGATATTATCGCCAAATATGATGTGTGTGCTACTGCGGCTGGAGTTCTTCTCT	558
Qy	141	ArgAspValIleThrValThrLeuProAspTyrArgIleSerValProIleProLeuThrVal	160
Db	559	CCTGATGTCACCGTACTCTCGCGGACTACCCCTGGTTCAAGGCCAAATTCCTTACCGCT	618
Qy	161	Tyr 161	
Db	619	TAT 621	
RESULT 9			
LOCUS	AX363712	903 bp	DNA linear PAT 15-FEB-2002
DEFINITION	AX363712	Sequence 207 from Patent EP1178052.	
ACCESSION	AX363712		
VERSION	AX363712.1	GI:18695826	
KEYWORDS			
SOURCE		Escherichia coli.	
ORGANISM		Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.	
FEATURES			
source		Location/Qualifiers	
		1..903	
		/organism="Escherichia coli"	
		/db_xref="taxon:562"	
BASE COUNT	206 a	213 c	240 g 244 t
ORIGIN			
Alignment Scores:			
Pred. No.:	1	0.3e-76	Length: 903
Score:		834.00	Matches: 158
Percent Similarity:		98.768	Conservative: 1
- Best Local Similarity:		98.144	Mismatches: 2
Query Match:		98.354	Indels: 0
DB:		6	Gaps: 0
US-09-900-575-29_COPY_26_186 (1-161) x AX363712 (1-903)			
Qy	1	ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis	20
Db	139	CCCGTCGCTGATGATGGGGGCAAAACCTGGTGTGGATCTTTCGACGAATCTTTGGCAT	198
Qy	21	AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgIleSerAlaTyrGly	40
Db	199	AACGATTATCCGGAACCATTAACGACTATGTCACTGCAACGAGCGCTGGCTTATGGC	258
Qy	41	GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro	60
Db	259	GGCGTGTATCTAATTTTTCGGGACCGGTAAATATGTGGCACTAGCTATCCATTTCTCT	318
Qy	61	ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal	80
Db	319	ACCACCGACGAAACCCGCGGCTGTTATATATCGAAGACGATTAAGCCGTGGCGCGTG	378
Qy	81	AlaLeuTyrIleThrProValSerSerAlaGlyIleValIleIleValAlaGlySerLeu	100
Db	379	GGCGTTTATTTAGACCGCTGTAGCAGTGGCGGCGGGGCGGATTAAGCTGGCTCATTA	438
Qy	101	IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal	120
Db	439	ATTGGCGGCTTATTGTCGACAGACCACAACTATAACACGATGATTTCCAGTTGTG	498
Qy	121	TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla	140
Db	499	TGGATATTATCGCCAAATATGATGTGTGTGCTACTGCGGCTGGAGTTCTTCTCT	558
Qy	141	ArgAspValIleThrValThrLeuProAspTyrArgIleSerValProIleProLeuThrVal	160

Db	559	CGTATGTCACCGTTACTGTCGCCGACTACCCCTGGTTCACTGCGCAATTCCTTACCGTT	618
Oy	161	Tyr 161	
Db	619	TAT 621	
RESULT 10			
LOCUS	AF154925	1040 bp	DNA linear BCT 10-JUL-2000
DEFINITION	Escherichia coli fimH precursor (fimH) gene, fimH-241 allele,		
ACCESSION	AF154925		
VERSION	AF154925.1	GI:5524627	
KEYWORDS	Escherichia coli.		
SOURCE	Escherichia coli.		
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.		
REFERENCE	1 (bases 1 to 1040)		
AUTHORS	Hamrick,T.S., Harris,S.L., Spears,P.A., Havelle,E.A., Horton,J.R., Russell,P.W. and Orndorff,P.E.		
TITLE	Genetic characterization of Escherichia coli type 1 pilus adhesin mutants and identification of a novel binding phenotype		
JOURNAL	J. Bacteriol. 182 (14), 4012-4021 (2000)		
MEDLINE	20373582		
PUBMED	10869080		
REFERENCE	2 (bases 1 to 1040)		
AUTHORS	Hamrick,T.S., Harris,S.L., Spears,P.A., Havelle,E.A., Horton,J.R., Russell,P.W. and Orndorff,P.E.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-MAY-1999) Microbiology, Pathology, and Parasitology, North Carolina State University, College of Veterinary Medicine, 4700 Hillsborough Street, Raleigh, NC 27606, USA		
FEATURES	Location/Qualifiers		
source	1..1040		
gene	/organism="Escherichia coli"		
	/strain="K12"		
	/db_xref="taxon:562"		
	47..949		
	/gene="fimH"		
	/note="mutant allele"		
	/allele="fimH-241"		
	47..949		
	/gene="fimH"		
	/note="type 1 pill adhesin"		
	/codon_start=1		
	/transl_table=11		
	/product="fimH precursor"		
	/protein_id="AAD4319.1"		
	/db_xref="GI:5524628"		
	/translation="MKRYITLFAVLIMGWSYNDMSFACKTRANGTAIPFGGSANVYVN LKPVVVGONLVVDLSTOIFCHNDVPETITDYPVLQKSAIGVLSNFSGTIVKSSGS YPEPTSETPRVYNSRTDKPWPVALYITPVSSAGVAIKGSLIAVLILKQNTNINS DDOEPFWNIYANNDDVYVPGGDSVARDVITLDPDGSYPIPLTVCAASQNGYYL SGTADAGNSITFTNTASFSPAGQGVQLTRNGTIIIPANNVTLSGAVGTSVAVSIGLTPAN YARTGQVTAGNVQSIIGVTIYQ"		
	102		
	/gene="fimH"		
	/note="confers an alanine to aspartic acid change near the signal sequence processing site"		
	/replace="c"		
BASE COUNT	253 a 243 c 270 g 274 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1,21e-76	Length:	1040
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	1	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AF154925 (1-1040)

```

QY      1  ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
Db      185  CCCGCTGTAATGTGGGGCAAAACCTGCTGTGATCTTTTCAGACGCAAAATCTTTGCCAT 244
QY      21  AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
Db      245  AACGATTATCCGGAACCATTTACGACTATGTCACACTGCACAGAGCTCGCTATGGC 304
QY      41  GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
Db      305  GCGCTGTAATCTAATTTTTCGGGACCGTAAATATAGTGGAGTAGTATCATCATTTCT 364
QY      61  ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProThrProVal 80
Db      365  ACCACGAGCAACGCCGCCGCTTGTATTAATTCAGAACGATTAAGCCGTCGGCGTG 424
QY      81  AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
Db      425  GCGCTTATTTTACGCGCTGTAGCACTGCGGCGCGGCTGCATTAAGCTGCTCATTA 484
QY      101  IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
Db      485  ATTGGCGTCTTATTTTGGCAGACACCACTAATACAGCGATGATTTCCAGTTTGTG 544
QY      121  ThrPasnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
Db      545  TGGAAATATTTACGCCAATATGATGTGGTGCTCTACTGCGGCGCTGGCATTTCTCT 604
QY      141  ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
Db      605  CGTGATGTACCGCTTACTCTGCGGAGTACCTGGTTGACGTGACCAATCTCTTACCGTT 664
QY      161  Tyr 161
Db      665  TAT 667

RESULT 11
AF154926      1040 bp      DNA      linear      BCT 10-JUL-2000
LOCUS      Escherichia coli f1mH precursor (f1mH) gene, f1mH-236 allele,
DEFINITION      complete cds.
ACCESSION      AF154926
VERSION      AF154926.1 GI:5524629
KEYWORDS
SOURCE
ORGANISM      Escherichia coli.
                Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
                Escherichia.
REFERENCE
AUTHORS      1 (bases 1 to 1040)
                Hamrick, T.S., Harris, S.L., Spears, P.A., Havel, E.A., Horton, J.R.,
                Russell, P.W. and Orndorff, P.E.
TITLE      Genetic characterization of Escherichia coli type 1 pilus adhesin
                mutants and identification of a novel binding phenotype
JOURNAL      J. Bacteriol. 182 (14), 4012-4021 (2000)
MEDLINE      20327582
PUBMED      10869080
REFERENCE
AUTHORS      2 (bases 1 to 1040)
                Hamrick, T.S., Harris, S.L., Spears, P.A., Havel, E.A., Horton, J.R.,
                Russell, P.W. and Orndorff, P.E.
TITLE      Direct Submission
JOURNAL      Submitted (28-MAY-1999) Microbiology, Pathology, and Parasitology,
                North Carolina State University, College of Veterinary Medicine,
                4700 Hillsborough Street, Raleigh, NC 27606, USA
FEATURES
source
                1..1040
                /organism="Escherichia coli"
                /strain="k12"
                /db_xref="taxon:562"
                47..949
                /gene="f1mH"
                /note="mutant allele"
gene

```

```

CDS
/allele="f1mH-236"
47..949
/gene="f1mH"
/note="type 1 pili adhesin"
/codon_start=1
/transl_table=11
/product="f1mH precursor"
/protein_id="AAB44320.1"
/db_xref="GI:5524630"
/translation="MKRVITLFAVLIMGNSVNAALPACTANGTAIPIGGSSANYVN
LAPVNVNGMLVVDLSTQLFCHNDYDETTIDYTLQROGSAAGVGLSNPSGTYKYGSS
YPPPTSEPRVYVNSRTDKPMFVALYLPVSAGCAVAKASLLAVLLRLQTNVNS
DDEQFVNIYANNDVVPVPGCDVSAKDVTYLPDPFGSVPLPLVYCAKSONLGYL
SGTTADGNSIFPNTASFSPAGVGVQLTRNGTILIPANTVSLGAVGSAVSLGLAN
VARTGGVYAGVAGNQSITIGVEFYQ"
108
/gene="f1mH"
/note="confers a serine to leucine change at the signal
sequence processing site"
/replacement="C"
BASE COUNT      252 a      243 c      270 g      275 t
ORIGIN
variation
Alignment Scores:
Pred. No.:      1.21e-76      Length:      1040
Score:      834.00      Matches:      158
Percent Similarity:      98.76%      Conservative:      1
Best Local Similarity:      98.14%      Mismatches:      2
Query Match:      98.35%      Indels:      0
DB:      1      Gaps:      0
US-09-900-575-29_COPY_26_186 (1-161) x AF154926 (1-1040)
QY      1  ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
Db      185  CCCGCTGTAATGTGGGGCAAAACCTGCTGTGATCTTTTCAGACGCAAAATCTTTGCCAT 244
QY      21  AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
Db      245  AACGATTATCCGGAACCATTTACGACTATGTCACACTGCACAGAGCTCGCTATGGC 304
QY      41  GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
Db      305  GCGCTGTAATCTAATTTTTCGGGACCGTAAATATAGTGGAGTAGTATCATCATTTCT 364
QY      61  ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProThrProVal 80
Db      365  ACCACGAGCAACGCCGCCGCTTGTATTAATTCAGAACGATTAAGCCGTCGGCGTG 424
QY      81  AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
Db      425  GCGCTTATTTTACGCGCTGTAGCACTGCGGCGCGGCTGCATTAAGCTGCTCATTA 484
QY      101  IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
Db      485  ATTGGCGTCTTATTTTTCGGCAGACACCACTAATACAGCGATGATTTCCAGTTTGTG 544
QY      121  ThrPasnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
Db      545  TGGAAATATTTACGCCAATATGATGTGGTGCTCTACTGCGGCGCTGGCATTTCTCT 604
QY      141  ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
Db      605  CGTGATGTACCGCTTACTCTGCGGAGTACCTGGTTGACGTGACCAATCTCTTACCGTT 664
QY      161  Tyr 161
Db      665  TAT 667

RESULT 12
AX357443      4637 bp      DNA      linear      PAT 13-FEB-2002
LOCUS      Sequence 167 from Patent WO0185208.
DEFINITION

```

ACCESSION AX357443
VERSION AX357443.1 GI:18674526
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Sebbel, P., Dunant, N., Bachmann, M., Tissot, A. and Lecheney, F.
TITLE Molecular antigen arrays and vaccines
JOURNAL Patent: WO 0185208-A 167 15-NOV-2001;
Cyos Biotechnology AG (CH); Sebbel, Peter (CH); Dunant, Nicolas
(CH); Bachmann, Martin (CH); Tissot, Alain (CH); Lecheney,
Franziska (CH)
FEATURES
source location/Qualifiers
1..4637
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="PFIMFGH"
BASE COUNT 1124 a 1156 c 1188 g 1169 t
ORIGIN
Alignment Scores:
Pred. No.: 6.74e-76 Length: 4637
Score: 834.00 Matches: 158
Percent Similarity: 98.76% Conservative: 1
Best Local Similarity: 98.14% Mismatches: 2
Query Match: 98.35% Indels: 0
Gaps: 0
DB: 6
US-09-900-575-29_COPY_26_186 (1-161) x AX357443 (1-4637)
QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 3441 CCGCGTGTGATGTCGGGCAAAACCTGTCGTCGATCTTTCAGCGCAAAVCTTTTGCCAT 3500
QY 21 AsnAspTyrProGlnThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 3501 AACGATTATCCGGAACCATTCACACTATGTACACTGCACAGAGCGTCGGCTTATGCG 3560
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
DB 3561 GCGGTGTATCTAATTTTCGGGACCGTAAATATAGTGGCAGTACATTCATTTCTCT 3620
QY 61 ThrThrSerGlnThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
DB 3621 ACCACCGAGAAACGCGCGGTGTTTATTAATTCAGAAAGGATTAAGCGTCGCCAT 3680
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
DB 3681 GCGCTTATTTGACGCTGAGCAGTGCAGCGGCGGTGCGATTTAAAGCTGCTCATTTA 3740
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnSerTyrAsnSerAspPheGlnPheVal 120
DB 3741 ATTGCGCGCTTATTTTGGCAGACCAACACTATTAACAGCATGATTTCCAGTTTGTG 3800
QY 121 TrpAsnIleTyrAlaAsnAsnSpValValValProThrGlyGlyCysAspValSerAla 140
DB 3801 TGGAAATATTTACGCCAATAAATGATGTGTGCTGCTACTGCGGCTCGATGTTTCTGCT 3860
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 3861 CGTGATGTACCGCTTACTCTCGGACATACCTGTTTCAAGTCCCAATTTCTTACCCTT 3920
QY 161 Tyr 161
DB 3921 TAT 3923
RESULT 13 7416 bp DNA linear PAT 06-FEB-2002
LOCUS AX354080
DEFINITION Sequence 46 from Patent WO2004496.
ACCESSION AX354080
VERSION AX354080.1 GI:18618994

KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Langermann, S., Revel, A., Auguste, C. and Burtin, J.
TITLE Fibin adhesin proteins and methods of use
JOURNAL Patent: WO 0204496-A 46 17-JAN-2002;
MEDIMUNE, INC. (US)
FEATURES
source location/Qualifiers
1..7416
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Sequence of plasmid pCGA126-1"
BASE COUNT 1858 a 1807 c 1908 g 1843 t
ORIGIN
Alignment Scores:
Pred. No.: 1.16e-75 Length: 7416
Score: 834.00 Matches: 158
Percent Similarity: 98.76% Conservative: 1
Best Local Similarity: 98.14% Mismatches: 2
Query Match: 98.35% Indels: 0
Gaps: 0
DB: 6
US-09-900-575-29_COPY_26_186 (1-161) x AX354080 (1-7416)
QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 1244 CCGCGTGTGATGTCGGGCAAAACCTGTCGTCGATCTTTCAGCGCAAAVCTTTTGCCAT 1303
QY 21 AsnAspTyrProGlnThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 1304 AACGATTATCCGGAACCATTCACACTATGTACACTGCACAGAGCGTCGGCTTATGCG 1363
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
DB 1364 GCGGTGTATCTAATTTTCGGGACCGTAAATATAGTGGCAGTACATTCATTTCTCT 1423
QY 61 ThrThrSerGlnThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
DB 1424 ACCACCGAGAAACGCGCGGTGTTTATTAATTCAGAAAGGATTAAGCGTCGCCAT 1483
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
DB 1484 GCGCTTATTTGACGCTGAGCAGTGCAGCGGCGGTGCGATTTAAAGCTGCTCATTTA 1543
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnSerTyrAsnSerAspPheGlnPheVal 120
DB 1544 ATTGCGCGCTTATTTTGGCAGACCAACACTATTAACAGCATGATTTCCAGTTTGTG 1603
QY 121 TrpAsnIleTyrAlaAsnAsnSpValValValProThrGlyGlyCysAspValSerAla 140
DB 1604 TGGAAATATTTACGCCAATAAATGATGTGTGCTGCTACTGCGGCTCGATGTTTCTGCT 1663
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 1664 CGTGATGTACCGCTTACTCTCGGACATACCTGTTTCAAGTCCCAATTTCTTACCCTT 1723
QY 161 Tyr 161
DB 1724 TAT 1726
RESULT 14 9299 bp DNA linear PAT 13-FEB-2002
LOCUS AX357444
DEFINITION Sequence 168 from Patent WO0185208.
ACCESSION AX357444
VERSION AX357444.1 GI:18674527
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.

REFERENCE 1
 AUTHORS Seibel, P., Dunant, N., Bachmann, M., Tissot, A. and Lechener, F.
 TITLE Molecular antigen arrays and vaccines
 JOURNAL Patent: WO 0185208-A 168 15-NOV-2001;
 Cytos Biotechnology AG (CH); Seibel, Peter (CH); Dunant, Nicolas
 (CH); Bachmann, Martin (CH); Tissot, Alain (CH); Lechener,
 Franziska (CH)
 FEATURES Location/Qualifiers
 source 1. 9299
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="PRIMAIDRGH"
 BASE COUNT 2371 a 2185 c 2410 g 2333 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.5e-75 Length: 9299
 Score: 834.00 Matches: 158
 Percent Similarity: 98.768 Conservative: 1
 Best Local Similarity: 98.148 Mismatches: 2
 Query Match: 98.354 Indels: 0
 Gaps: 0
 US-09-900-575-29_copy_26_186 (1-161) x AX357444 (1-9299)
 Oy 1 ProvValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
 Db 8109 CCGGTGCTGATGATGCGGCAAAACCGCGTGGATCTTCAGCCAAATCTTTGCCAT 8168
 Oy 21 AsnAspTyrProGlnThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
 Db 8169 AACGATATCCGGAACCTTACAGACTATGTGCACATGCACAGAGCGCTGCTATATGCG 8228
 Oy 41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
 Db 8229 GCGGTGATCTTAATTTTCCGCGACCGTAAATATAGCGAGTATCCATTTCTT 8288
 Oy 61 ThrThrSerGlnThrProArgValValLysAsnSerArgThrAspLysProTyrProVal 80
 Db 8289 ACCACACGAGAAACCGCGCGCTTTATATTCGACAAACGATAACCGCTGCGCGTG 8348
 Oy 81 AlaLeuTyrLeuThrProValSerAlaGlyLeuValIleLysAlaGlySerLeu 100
 Db 8349 GCGCTTTATTTGACGCCCTGTGAGCAGTGGCGGCGGCGGATTAAGCTGCTCATTA 8408
 Oy 101 IleAlaValIleuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
 Db 8409 ATTGCGGCTTATTTTGGACAGACCAACATATTAACCGCATTTCCACTTTGTG 8468
 Oy 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
 Db 8469 TGGAAATATTTACGCCAATATATGATGTGTGCTGCTGCGGCTCGATGTTTCTGCT 8528
 Oy 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIlePheLeuThrVal 160
 Db 8529 CGTATGTCACCGTACTCTGCGGACTACCGTCACTGTCATTCCTCTTACCGTT 8588
 Oy 161 Tyr 161
 Db 8589 TAT 8591
 RESULT 15
 LOCUS AE000502 11313 bp DNA linear BCT 01-DEC-2000
 DEFINITION Escherichia coli K12 MG1655 section 392 of 400 of the complete
 genome.
 ACCESSION AE000502 UN00096
 VERSION AE000502.1 GI:2367374
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Escherichia coli K12.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1
 AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Berna, N.T., Burland, V.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mau, B. and Sho, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503
 REFERENCE 2 (bases 1 to 11313)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 REFERENCE 3 (bases 1 to 11313)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 REFERENCE 4 (bases 1 to 11313)
 AUTHORS Plunkett, G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 COMMENT On Sep 9, 1997 this sequence version replaced 91.1790766.
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
 30332 (e-mail: mark@ember.gatech.edu). Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>). ** The E. coli K12 sequence and
 its annotations are periodically updated: this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'p' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.
 FEATURES
 source 1. 11313
 Location/Qualifiers
 1. 11313
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 429..443
 /note="central position to flmB promoter: -211.5"
 /bound_moiety="Lrp documented site"
 613..642
 /note="Factor Sigma70; predicted +1 start at 4538423"
 751..1353
 /gene="flmB"
 /gene="flmB"
 /note="b4312"
 751..1353
 /gene="flmB"
 /function="regulator; Surface structures"

```
/note="o200; 100 pct identical to FIMB_ECOLI SW: P04742;  
CG Site No. 18355; alternate gene name pil"  
/codon_start=1  
/transl_table=11  
/product="recombinase involved in phase variation;  
regulator for fimA"  
/protein_id="AAC77268.1"  
/db_xref="GI:1790767"  
/translation="MKKADNKKRNPLTHSEITSLIKANTGPHARNYCLTLLCFIH  
GFRASEICRLISDIDLAKACIYIHLKGFSTTHPLKNENQALKNWSISIRSYHA  
ESEWFLSRKGNPLSRQFYHISTSGNGLSLHPHMLRSCFPALANMGIDTRL  
IODYLRNIRHFWYTSNAGRFGYGIWDRGRORRAVL"  
1831..2427  
/gene="fimB"  
/note="b4313"  
1831..2427  
/gene="fimB"  
/function="regulator: Surface structures"  
/note="o198; 100 pct identical to FIME_ECOLI SW: P04741;  
CG Site No. 18346"  
/codon_start=1  
/transl_table=11  
/product="recombinase involved in phase variation;  
regulator for fimA"  
/protein_id="AAC77269.1"  
/db_xref="GI:1790768"  
/translation="MSKRRYLTKKEVQAMQAVCYGATGARDYCLILAYRHGMRIS  
LIDYIOLDNIEGRINIRLKNESYTHPLFDEREVERWTOERANMKGADRTDL  
LISRGSRSLSRQARITRIDGIEAGTYOTPHMLRACGYELAEAGDRTLIDYLL  
GHRNIRHRYTSNAGRFGIEMERNNILNKLKREY"  
2909..3457  
/gene="fimA"  
/note="b4314"  
2909..3457  
/gene="fimA"  
/function="structural component: Surface structures"  
/note="o182; 100 pct identical to FIML_ECOLI SW: P04128;  
CG Site No. 18358; alternate gene names fimD, pilA"  
/codon_start=1  
/transl_table=11  
/product="major type 1 subunit fimbriae (pilin)"  
/protein_id="AAC77270.1"  
/db_xref="GI:1790769"  
/translation="MKIKLAIYVLSAISTSTALAAATVNGGTVHFKRQVNAAC  
AVDAGSDQTVLQGVRTASLAEQATSAVFNFQINDCPNNAKAAVATLGAID  
AGHTNVALQSSAAGSATVGVQIILDTGAULTDGAFTSSETTLNNGTINTPQARY  
FATGATPGAAANDATFKVOYQ"  
3414..4061  
/gene="fimI"  
/note="b4315"  
3414..4061  
/gene="fimI"  
/function="structural component: Surface structures"  
/note="o215"  
/codon_start=1  
/transl_table=11  
/product="fimbrial protein"  
/protein_id="AAC77271.1"  
/db_xref="GI:1790770"  
/translation="MLMRMRPSRFSINNLPREFDITGDHAFCAIKITTKRRRLFL  
LASLLPMEALAGNKWNTLLPGNMQFQVIAETCRLEADDKQMTNMGQISSNRHA  
VGEPSAVPEVILHREGSTVSESVGFVAFHGDGKMPDLVSGEGGIAATNINRALF  
DDGGLNVPINRPRANKRKLXSGTSLHFIKRYATGRYTGIAANQAMPSLTYQ"  
4098..4823  
/gene="fimC"  
/note="b4316"  
4098..4823  
/gene="fimC"  
/function="factor: Surface structures"  
/note="o241; 100 pct identical to FIMC_ECOLI SW: P31697;  
CG Site No. 18352; alternate gene name pil"  
/codon_start=1  
/transl_table=11
```

```
/product="periplasmic chaperrone, required for type 1  
fimbriae"  
/protein_id="AAC77272.1"  
/db_xref="GI:1790771"  
/translation="MSKNKNVRSQETFPCLAGLIMFAMNAYAGRAEAGVALGATR  
VIVPAGQEOELAVTNDENSTYLIQSVENADGVKGRFTVPELPAMGKENTLR  
ILDANNLPQDRESLEPMNAYKALIPMSKSTLENTQLA1ISKIKLYPALUALPP  
DOAKELRRRSANSLTILNTPYLLVTELNASTRLVLENALVPMGESVTKLPSDAG  
SNITRTINDYGALTRKMTGYME"  
4891..7527  
/gene="fimD"  
/note="b4317"  
4891..7527  
/gene="fimD"  
/function="membrane: Outer membrane constituents"  
/note="o878; 100 pct identical to FIMD_ECOLI SW: P31030;  
CG Site No. 18349"  
/codon_start=1  
/transl_table=11  
/product="outer membrane protein: export and assembly of  
type 1 fimbriae, interrupted"  
/protein_id="AAC77273.1"  
/db_xref="GI:1790772"  
/translation="MSYLNRLYQNTQCLHTRKRLAGFPYRLVACAFAPAOPLSS  
ADLYENRFLADPOAVADLSRFENGQELPETHYVDIYLLNGYMATPDVTPNGDSE  
OGIVPCLTRRAQLSMGLNTASVAGNMLADDCVPLTWVQDATHLDVGOQLNLT  
POAFMSNRARGVTPPELMDPGINAGLLNTEGNSVQNTGNSHAYLNLQSGNLIG  
AMRLRNTMYSVNSLSDRSSGSKNMKQHINTWLEBRIIPKRSLLTGDYTGQDIFDGI  
NFRGADLSDNNMLPDSORGSPAPVHIGIRGAOYTIKONGDIANTVPPPTIND  
IVAGNSGDLQVYTRKADSDTOIFTPYVSSVPLDREGHRTYSITAGEVRSQAQEK  
TRFQOSTLHLGAPAGWTITGGTOLADRYRANPFGICKNGALISVDMTOASTLPD  
DSOQSDQSVRELYNKSLSNGINQLVGRIRYTSSTGIVLSSHOTYSGTSMHARSYSMS  
IYKPFETDYNNLAYNKRGLQLTIVQJGRSTLYLSSHOTYSGTSMHARSYSMS  
NTAFEDINMTLSYSLTKNWKQKQDMLANVNIPIHMLRSDSSQMHARSYSMS  
HDLNGRMTNLAGYVGLLEDNMLNLSYVQFVAGGGSSTGYATLNRGGYGNAT  
GYSHSDIRQLYVYGGSVLYANANGYTLQPLNDYVLVAKGAPDAKYNQTVGRTD  
WRGTAIVPIATETRENRNVALDTTTLADNVDNAVNVNPGATVIRAFKARVGIKL  
LMTLTNNKRLPGAAVTSSESSSGSVADVANGVYLSGMPLAGKQVQKGEENAHCV  
ANYQLPESQOQLTQLSACR"  
7537..8067  
/gene="fimF"  
/note="b4318"  
7537..8067  
/gene="fimF"
```

Alignment Scores:

Pred. No.:	1.88e-75	Length:	11313
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	1	Gaps:	0

US-09-900-575-29_copy_26_186 (1-161) x AE000502 (1-11313)

QY	1	ProvaIvalAsnVAlGlyInsnLeuValIaAspLeuSerThgInIlePheCysHis	20
DB	8741	CCCGTGTGAATGTGGGCAAAACCTGCTGATCTTTCAGCCCAATCTTTGGCAT	8800
QY	21	AsnAspTyProglunThrIleThrAspTyValThrLeuGlnArgGlySerAlaTyrlly	40
DB	8801	AACGATTATCCGGAACCAATTCACAGACTATGCACACTCCACAGAGCGCTGCTTATG6C	8860
QY	41	GlyValIeuSerAsnPheserGlyThrValLysTySerGlySerSerTyProPhePro	60
DB	8861	GGCGTGTATCTAATTTTCGGGACCGTAATAATATAGTGGCAGTATCATTTCT	8920
QY	61	ThrThrSerGluThrProArgValValTyAsnSerArgThrAspLysProTrProval	80
DB	8921	ACCACCGCAAAACCGCGCTGTTTATTAATTGAGAACGAGATTAACCGCGCGGTG	8980
QY	81	AlaLeuTyrlleuThrProvalSerSerAlaGlyLeuValIleLysAlaGlySerLeu	100

gene /note="predicted bend of 72.06 degrees"
4279. .4728
/gene="y1cA"
4279. .4728
/gene="y1cA"
/codon_start=1
/transl_table=1
/label=ORF_0149
/protein_id="AA96989.1"
/db_xref="GI:536934"
/translation="MKRIAFCDHVGEPILKHEIVAHVGVGVVIDKGTWSSERTDYP
HYASQVALAVAGGEVDGILICGTGISTIAANKPACIRAVYVSEPSQSLSRQNDHT
NYLAFGSRVVGLELAKMIVDMLGADYEGGRHOORVEALITAIQRKN"
complement(4520. .16191)
/note="corresponds to Genbank Accession Number D90227
(ECOPHN) 1. .11672: the related sequence Genbank Accession
Number J05260 (ECOPHNAO) is from E. coli B, not K-12"
misc_feature complement(4739. .5146)
gene /gene="phnQ"
complement(4739. .5146)
/gene="phnQ"
/note="annotated in Genbank Accession Number J05260
(ECOPHNAO), but probably not a gene"
/codon_start=1
/transl_table=1
/protein_id="AA96990.1"
/db_xref="GI:536935"
/translation="MPAPYPTSPILIFHSPLFLQPLTLIKIALAIYOLPAIEKL
AAAYILSAPIVAHFTLPAAVARIFIVSPYRVISVNLINTANGVVDNIGLRND
SVSRVKNKSGKRFISOPKETTSLITITHPCK"
complement(5272. .6030)
gene /gene="phnB"
complement(5272. .6030)
/gene="phnB"
/codon_start=1
/transl_table=1
/protein_id="AA96991.1"
/db_xref="GI:536936"
/translation="MSLTLTLTGAGAGVPMGCECAACARRSPQYRPPCSGVV
KFNDAITLIDAGLHDLADKMSFGSQOPLITHYMDHYGSPILRMCGDDIPYGGP
DEQGGDDLFEKHGLIDFSHTVEPVVDLQGLQVPLPLNNSKITLFTGLLETSRVA
WLSDPGPEKTLKFLRNNOQVVMDCSHPPRADAPRNHCDLNTLVLANOVTSRPRV
ILTIHSQEDALMENALPSGFVEGFDMEIGVA"
misc_structure 5452. .5586
/note="predicted bend of 75.60 degrees"
complement(6032. .6466)
gene /gene="phnO"
complement(6032. .6466)
/gene="phnO"
/codon_start=1
/transl_table=1
/protein_id="AA96992.1"
/db_xref="GI:536937"
/translation="MPACELRATQYDDDAVYALICELKQAEFHDAFRVGNANLRD
PMRYHALDELVEVGMIGLILQFHLHNNHNGEIOELVYMPQAGLNVGSLLMWAE
EERARQAGEMTELSNVKRNDAHRYLRGVBQSHRETKAL"
complement(6453. .7010)
gene /gene="phnN"
complement(6453. .7010)
/gene="phnN"
/note="HisP-like nucleotide binding protein (ATP-binding
protein)"
/codon_start=1
/transl_table=1
/protein_id="AA96993.1"
/db_xref="GI:536938"
/translation="MMGKILIMGPSSGSKDSLAEILRLDEQTLIVAHRYTIRDASA
GSENNHLSQEFETRAQONLLASWHAGLTYGVGVETDLMHAGFDVLYGSRHL
PQARAYOSALPLVCLQVSPETIKORLEKRGRENASEINARLARARYTPQDCHTLNN
DGLRSQVDFLLTIHQEKHHACL"
complement(7010. .8146)
gene /gene="phnM"
complement(7010. .8146)

/gene="phnM"
Alignment Scores:
Pred. No.: 9,42e-74 Length: 338534
Score: 834.00 Matches: 158
Percent Similarity: 98.76% Conservative: 1
Best local Similarity: 98.14% Mismatches: 2
Query Match: 98.35% Indels: 0
Gaps: 0
US-09-900-575-29_copy_26_186 (1-161) x EC00W93 (1-338534)
QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrcGlnIlePheCysHis 20
|||||
Db 239779 CCCGTCGTGATGTGGGCAAAACCTGCGTGCATCTTCGACGCAATCTTTGCCAT 239838
|||||
QY 21 AsnAspTyrProGluThrIleThrAspTyrValIleLeuGlnArgGlyGlnAlaTyrGly 40
|||||
Db 239839 AACGATATCCGGAACCACTACGACTATGTCACACTGCACAGAGGCTGGCTTATGCG 239898
|||||
QY 41 GlyValLeuSerAsnPheserGlyThrValIleTyrSerGlySerTyrProPhePro 60
|||||
Db 239899 GGCGTGTATCTATTTTCCGGGACCGTAAATATAGTGCAGTATCATTCATTCCT 239958
|||||
QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTTPProVal 80
|||||
Db 239959 ACCACGACGGAACGCCGCCGTTGTTATATTCGGAACGGATTAAGCCGTCGCGTG 240018
|||||
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
|||||
Db 240019 GCGCTTATTTGACGCTGTGAGCAGTGCAGGCGGGTGGCGCATTAAGCTGGCTCATTA 240078
|||||
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
|||||
Db 240079 ATTCGCGTGTATTTTGGCAGACCAACCACTATTAACGCGTGAATTTCCAGTTGTG 240138
|||||
QY 121 TrpAsnIleTyrAlaAsnAspValValValProThrGlyGlyCysAspValSerAla 140
|||||
Db 240139 TGGATATTTTGGCCAAATAAGATGTGTGTGCTACTGCGGCTCGATGTTTCGCT 240198
|||||
QY 141 ArgAspValThrValIleThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
|||||
Db 240199 CGTGATGTACACGTTTACCTCGCGAGTACCTGTTACAGTGCACAAATTCCTTACCGTT 240258
|||||
QY 161 Tyr 161
|||
Db 240259 TAT 240261
RESUT 17
AX354088
LOCUS AX354088 837 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 54 from Patent WO0204496.
ACCESSION AX354088
VERSION AX354088.1 GI:18619002
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 Langemann,S., Revel,A., Auguste,C. and Buellein,J.
AUTHORS Film adhesin proteins and methods of use
TITLE Patent: WO 0204496-A 54 17-JAN-2002;
JOURNAL MEDIMUNE, INC. (US)
FEATURES
source location/Qualifiers
1. 837
/organism="Escherichia coli"
BASE COUNT 192 a 201 c 221 g 223 t
ORIGIN
Alignment Scores:
Pred. No.: 1.91e-76 Length: 837

Score: 831.00 Matches: 157
 Percent Similarity: 98.76% Conservative: 2
 Best Local Similarity: 97.52% Mismatches: 2
 Query Match: 98.00% Indels: 0
 DB: 6 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AX354088 (1-837)

QY 1 ProValValAsnValAGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
 DB 76 CCGCGTGAATGTGGGCAAAACCTGCTGCGATCTTTGAGCGCAATCTTTGGCAT 135
 QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
 DB 136 AACGATTATCCGGAACCATTTACAGACTATGACACTGCACAGAGCGCTGCGTTATGCG 195
 QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyr-ProPhePro 60
 DB 196 GCGGCTTATCTAATTTTCCGGACCGTAAATATATATGACAGTACATCATTTCCCT 255
 QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
 DB 256 ACCACGACGCAACGCCCGCGGTGTATATTCGACAGACGATMAAGCCGCGCGTG 315
 QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
 DB 316 GCGCTTATTTGACGCGCTGTGACAGAGTGGCGGCGGCGCATTTAAAGCTGGCTCATTTA 375
 QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAspTyrAsnSerAspAspPheGlnPheVal 120
 DB 376 ATTGCGGTATTTATTTTGGCAGACACCAACATATTAACGCGATTTTCCAGTTTGTG 435
 QY 121 TrpAsnIleTyrAlaAsnAspValValAlaProThrGlyGlyCysAspValSerAla 140
 DB 436 TGGATATTTTACGCCAATAATGATGTGTGCTACTGGCGCTGCGATGTTTCTGCT 495
 QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
 DB 496 CGTATGTACCGTACTGTGCGGACTACCGCTGTTCAATGCCAATTCCTTACCGTT 555
 QY 161 Tyr 161
 DB 556 TAT 558

RESULT 18
 AX354037
 LOCUS AX354037 837 bp DNA linear PAT 06-FEB-2002
 DEFINITION Sequence 3 from Patent WO0204496.
 ACCESSION AX354037
 VERSION AX354037.1 GI:18618974
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE
 AUTHORS Langermann, S., Revel, A., Auguste, C. and Burteln, J.
 TITLE Film adhesin proteins and methods of use
 JOURNAL Patent: WO 0204496-A 3 17-JAN-2002;
 MEDIMUNE, INC. (US)
 FEATURES
 source 1..837
 Location/Qualifiers

BASE COUNT 193 a 201 c 221 g 222 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2.43e-76 Length: 837
 Score: 830.00 Matches: 157
 Percent Similarity: 98.14% Conservative: 3
 Best Local Similarity: 97.52% Mismatches: 1
 Query Match: 97.88% Indels: 0

DB: 6 Gaps: 0
 US-09-900-575-29_COPY_26_186 (1-161) x AX354037 (1-837)

QY 1 ProValValAsnValAGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
 DB 76 CCGCGTGAATGTGGGCAAAACCTGCTGCGATCTTTGAGCGCAATCTTTGGCAT 135
 QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
 DB 136 AACGATTATCCGGAACCATTTACAGACTATGACACTGCACAGAGCGCTGCGTTATGCG 195
 QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyr-ProPhePro 60
 DB 196 GCGGCTTATCTAATTTTCCGGACCGTAAATATATGTCGAGTACATCATTTCCG 255
 QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
 DB 256 ACCACGACGCAACGCCCGCGGTGTATATTCGACAGACGATMAAGCCGCGCGTG 315
 QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
 DB 316 GCGCTTATTTGACGCGCTGTGACAGAGTGGCGGCGGCGCATTTAAAGCTGGCTCATTTA 375
 QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAspTyrAsnSerAspAspPheGlnPheVal 120
 DB 376 ATTGCGGTATTTTGGCAGACACCAACATATTAACGCGATTTTCCAGTTTGTG 435
 QY 121 TrpAsnIleTyrAlaAsnAspValValAlaProThrGlyGlyCysAspValSerAla 140
 DB 436 TGGATATTTTACGCCAATAATGATGTGTGCTACTGGCGCTGCGATGTTTCTGCT 495
 QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
 DB 496 CGTATGTACCGTACTGTGCGGACTACCGCTGTTCAATGCCAATTCCTTACCGTT 555
 QY 161 Tyr 161
 DB 556 TAT 558

RESULT 19
 AF154928
 LOCUS AF154928 1040 bp DNA linear BCT 10-JUL-2000
 DEFINITION Escherichia coli film precursor (film) gene, film-205 allele,
 complete cds.
 ACCESSION AF154928
 VERSION AF154928.1 GI:5524633
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE
 AUTHORS 1 (bases 1 to 1040)
 Hamrick, T.S., Harris, S.L., Spears, P.A., Havel, E.A., Horton, J.R.,
 Russell, P.W. and Orndorff, P.E.
 TITLE Genetic characterization of Escherichia coli type 1 pilus adhesin
 mutants and identification of a novel binding phenotype
 JOURNAL J. Bacteriol. 182 (14), 4012-4021 (2000)
 MEDLINE 20272582
 PUBMED 10869080

REFERENCE
 AUTHORS 2 (bases 1 to 1040)
 Hamrick, T.S., Harris, S.L., Spears, P.A., Havel, E.A., Horton, J.R.,
 Russell, P.W. and Orndorff, P.E.
 TITLE Direct Submission
 JOURNAL Submitted (28-MAY-1999) Microbiology, Pathology, and Parasitology,
 North Carolina State University, College of Veterinary Medicine,
 4700 Hillsborough Street, Raleigh, NC 27606, USA
 FEATURES
 source 1..1040
 Location/Qualifiers

1..1040
 /organism="Escherichia coli"
 /strain="K12"
 /db_xref="taxon:562"
 gene 47..949

OY	61	ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspIysProTyrProVal	80
Db	256	ACCACACCGAAGACCCCGCGCTGTATTATATTCGAAACGGAATAGCCCTGGCCGGTG	319
OY	81	AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleIysAlaGlySerLeu	100
Db	316	GCGCTTATTATTCAGCGCCCTGTGAGCAGTGCGGGGGGGCGCGGATTAAGCGCTCATTTA	375
OY	101	IleAlaValIleuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal	120
Db	376	ATTGCCGGCTTATTATTTCGCAACACCAACCACTATAACACGATGATTTCCAGTTGTG	435
OY	121	TrpAsnIleTyrAlaAsnAsnAspValValValProthrGlyGlyCysAspValSerAla	140
Db	436	TGGATATTTCAGCCCAATATGATGTGTGTGTGCTCTAGCGGCTCCGATGTTTGTCT	495
OY	141	ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal	160
Db	496	CGTGATGTACACCGTTACTCTCCGGACCTACCCGTGGTTCAGTGGCAATTCCTCTTACCGTT	555
OY	161	Tyr 161	
Db	556	TAT 558	

DEFINITION	Escherichia coli F1mH precursor (f1mH) gene, f1mH-166 allele.
ACCESSION	AF306536.1
VERSION	GI:1120563
KEYWORDS	Complete cds.
ORGANISM	Escherichia coli.
SOURCE	Escherichia coli.

REFERENCE AUTHORS	TITLE	JOURNAL	MEDLINE
1 (bases, 1 to 1040) Harris,S.L., Spears,P.A., Haveli,E.A., Hamrick,T.S., Horton,J.R. and Orndorff,P.E.	Characterization of <i>Escherichia coli</i> type 1 plus mutants with altered binding specificities	J. Bacteriol.	183 (13), 4099-4102 (2001) 21289098

FORMED	11352476
REFERENCE	2 (bases 1 to 1040)

TITLE Direct Submission
JOURNAL Submitted (19-SEP-2000) Microbiology, Pathology and Parasitology

```

FEATURES
4700 Hillsborough Street, Raleigh, NC 27606, USA
source
1. 1040
Location/Qualifiers
/organism="Escherichia coli"
/strain="K12"
/db_xref="taxon:562"
47. 949
gene
```

47:349
/gene="flmH"
/note="type 1 pili adhesin; mutation confers a
phenylalanine to serine change"
/codon_start=1
/transl_table=1
/product="flmH precursor"
/protein_id="ACG0926.1"
/db_xref="GI:1120564"
/translation="MKRVTLEFVLMLGKSNVAFACKTANGTAIPITGGSANVYNN
LAPVTGONLIVYDLSOLFCHNDYPERITDVTLRGSAVCVSNISGTVKISGSS
YLPPTTEPTGVRNSTRKQKMPVAILTRDVSACGVKIKOST

variation
534

BASE COUNT	252 a	245 c	270 g	273 t
ORIGIN				

Alignment Scores:	
Pred. No.:	8, 04e-76
Score:	826.00
Percent Similarity:	98.14%
Best Local Similarity:	97.52%
Query Match:	97.41%
DB:	1
	0
Length:	1040
Matches:	157
Conservative:	1
Mismatches:	3
Indels:	0
Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AF306536 (1-1040)

OY	1	ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheGlyHis	20
Db	185	CCCGCGGAAATGGGGGCAAAACCTGGTGGATCTTTCGACGGCAATCTTTGGCAAT	244
OY	21	AsnAspYyrProGluThrIleThrAspYyrValThrLeuGlnArgGlySerAlaIArgIly	40
Db	245	AACGATTATCCGGAAACCATTTACGACTAGTGCACACTGCACGAGAGCTCGGCTTATGGC	304
OY	41	GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerYyrProPhePro	60
Db	305	GCCGCGTATTCTAATTTTTCCGGACCGCTAAATAATGTGGAGTATGCTATTCATTCTCT	364
OY	61	ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal	80
Db	365	ACCACACGCAAGCCGCCGCGTGTATTATATTCGAGAACGATAAAGCCGTGGCGGTG	424
OY	81	AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu	100
Db	425	GCGCTTATTTCAGCCCTGTGACGACATGGCGGCGGTGGCGATTAAAGCTGGCTATTAA	484
OY	101	IleAlaValLeuIleLeuArgGlnThrAsnAspYyrAsnSerAspPheGlnPheVal	120
Db	485	ATTGGCGGCTATTATTTTCGACAGACCAACACTATTAACACCGATGATTTCCACGTTTGTG	544
OY	121	TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla	140
Db	545	TGCATATTATTAAGCCAAATAAATGATGTGTGTGCTGCCTATGGCGGCTCGCATGTTTCGCT	604
OY	141	ArgAspValThrValThrLeuProAspYyrArgGlySerValProIleProIleuThrVal	160
Db	605	CTGTGATGTCACCGTATCTGCGGACATCACTGCTTTCAGGCCAATTCCTTTACCGGTT	664

RESULT 23			
AX354039			
LOCUS	AX354039	840 bp	DNA
DEFINITION	Sequence 5 from Patent WO0204496.		linear
ACCESSION	AX354039		
VERSION	AX354039.1	GI:18618976	

```

REFERENCE
1
AUTHORS
Langermann,S., Revel,A., Auguste,C. and Burelajn,J
TITLE
Fimh adhesin proteins and methods of use
JOURNAL
Patent: WO 0204496-A 5 17-JAN-2002;
MEDLINE
MEDIMUNE, INC. (US)
FEATURES
source
1. 840
/organism="Escherichia coli"

```

```
BASE COUNT      196 a      197 c      222 g      225 t
ORIGIN          /db_xref="taxon:562"
```

Alignment Scores:	
Pred. No.:	7,97e-76
Score:	825.00
Percent Similarity:	97.52%
Best Local Similarity:	96.88%
Query Match:	97.23%
DB:	6
Gaps:	0
Length:	8
Matches:	1
Conservative:	1
Mismatches:	4
Indels:	0
Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) X AX354039 (1-840)

QY	1	ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis	20
Db	76	CCTGCCGGGAATGTGGGGCAAAACCTGGTGTGATCTTTTCGACGCAGAAATCTTTGCCAT	135
QY	21	AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgIleSerAlaTyrGly	40
Db	136	AMCGATTTCGGGAACCATTCAGACTATGTCACTGCCACGAGGCTCGGCTATGGC	195
QY	41	GlyValIleuSerAspSerGlyThrValLysThrSerGlySerTyrProPhePro	60
Db	196	GGCGGTATATCAATTTTTCGGGACCGTAAATATATGTGGCAGTACATTCATTTCCG	255
QY	61	ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProThrProVal	80
Db	256	ACTACCAGGAAACCCCGCGGGTGTATTATTCGGAACGGGTAAAGCCGTGGCCGGTG	315
QY	81	AlaIleuTyrLeuThrProValSerSerAlaGlyIleValIleValAlaGlySerLeu	100
Db	316	GCGCTTTATTATTCACCCCTGTGACAGTCGGGTGGGGGGAGATTAAAGCTGGCTCATTA	375
QY	101	IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnIlePheVal	120
Db	376	ATTGGCCGTCTATTATTTCGCAGACGCCAACCACTAATACACGATGATTTCCAGTTTGTC	435
QY	121	TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla	140
Db	436	TGGAAATTTTACGCCAATATATGTGTGGTGCTACTGCGGCTCTCGATGTTTCTGCT	495
QY	141	ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal	160
Db	496	CATGATGTCCACCGTTACTCTGCCGAGTCACTCGTGTACGTCCAAATTCCTTTACCGTT	555
QY	161	Tyr 161	
Db	556	TAT 558	

RESULT	24				
AX354054					
LOCUS	837 bp	DNA	linear	PAT 06-FEB-2002	
DEFINITION	Sequence 20 from Patent WO0204496.				
ACCESSION	AX354054				
VERSION	AX354054.1				
KEYWORDS	GI:18618991				
SOURCE	.				
ORGANISM	Escherichia coli.				
	Escherichia coli.				
	Bacterii; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
	Escherichia				

REFERENCE	1	Langermann, S., Revel, A., Auguste, C. and Burelin, J
AUTHORS		
TITLE		Ftmh adhesion proteins and methods of use
JOURNAL		Patent: WO 020496-A 20 17-JAN-2002;
		MEDIMMUNE, INC. (US)
FEATURES		
source		Location/Qualifiers
		1..837
		/organism="Escherichia coli"
		/db_xref="taxon:562"
BASE COUNT	196 a	221 t
ORIGIN	200 c	220 g

BASE COUNT ORIGIN	196 a	200 c	220 g	221 t
----------------------	-------	-------	-------	-------

Alignment Scores:

Pred. No.: 1.01e-75 Length: 837
 Score: 824.00 Matches: 156
 Percent Similarity: 97.52% Conservative: 1
 Best Local Similarity: 96.89% Mismatches: 4
 Query Match: 97.17% Indels: 0
 DB: 6 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AX354054 (1-837)

Oy 1 ProValValAspValGlyGlnAsnLeuValAlaSpleSerThrGlnIlePheCysHis 20
 Db 76 CCGTCCGCGTGAATGGGCAAAACCTGGTGTAGATCTTTGGACGCAATCTTTGGCCAT 135
 Oy 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
 Db 136 AACGATTTATCCGGAACCATTTACAGACATATGTACACAGCAAGAGCTCGCTTATGGC 195
 Oy 41 GlyValLeuSerAsnPheSerGlyThrValIleTyrSerGlySerTyrProPhePro 60
 Db 196 GCGGTATATCTAATTTTCCGGACCGCTAAATATAGTGCAGTACATCATTTCCG 255
 Oy 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProThrProVal 80
 Db 256 ACCACACGCAACGCGCGCGGTGTATATTCGAGAACGATTAAGCGTGGCGCGTG 315
 Oy 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
 Db 316 GCGCTTATTTGACGCGCTGTGACAGAGTGGCGGGGTGGCGATTAAAGCTGCTCATTA 375
 Oy 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
 Db 376 ATTCGCGGTATATTTGGCGACAGCAAAACATTAACGCGATTCATTCAGATTGTG 435
 Oy 121 TrpAsnIleTyrAlaAsnAspValValValProThrGlyGlyCysAspValSerAla 140
 Db 436 TGCATATTTACGCCCATTAATGATGTGTAGTGTCTCTGCGCGCTGCGATGTTCTGCT 495
 Oy 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
 Db 496 CGTATATGTACCGTTACTCTGCGGACTACCGTTCAGTCCCAATTCTTACCGTT 555
 Oy 161 Tyr 161
 Db 556 TAT 558

RESULT 25

AE005662

LOCUS

Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 281

DEFINITION

AE005662

ACCESSION

AE005662

VERSION

AE005662.1

KEYWORDS

GI:12519313

SOURCE

Escherichia coli O157:H7 EDL933.

ORGANISM

Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae: Escherichia.

REFERENCE

1 (bases 1 to 14466)

AUTHORS

Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
 Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
 Postal, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
 Grobeck, E.J., Davis, N.W., Llin, A., Dimalanta, E., Potamoukis, K.,
 Apodaca, J., Anantharaman, T.S., Llin, J., Yen, G., Schwartz, D.C.,
 Welch, R.A. and Blattner, F.R.
 Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
 Nature 409 (6819), 529-533 (2001)
 21074935
 11206551

Grobeck, E.J., Davis, N.W., Llin, A., Dimalanta, E., Potamoukis, K.,
 Apodaca, J., Anantharaman, T.S., Llin, J., Yen, G., Schwartz, D.C.,
 Welch, R.A. and Blattner, F.R.
 Direct Submission
 Submitted (22-OCT-2000) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 Location/Qualifiers
 1. 14466
 /organism="Escherichia coli O157:H7 EDL933"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="enterohaemorrhagic"
 <1. 1267
 /note="O-157: Region of the EDL933 chromosome not
 homologous to E. coli K-12 MG1655"
 235. 267
 /gene="25903"
 235. 267
 /gene="25903"
 /function="orf; unknown function"
 /note="No significant matches"
 /codon_start=1
 /transl_table=11
 /protein_id="AA059489.1"
 /db_xref="GI:12519314"
 /translation="MIDILQVVCQ"
 complement(730..975)
 /gene="25904"
 complement(730..975)
 /gene="25904"
 complement(730..975)
 /function="orf; unknown function"
 /note="Residues 18 to 77 of 81 are 94.99 pct identical to
 residues 11 to 70 of 196 from Genpept 118 :
 g11787491|gb|AAC74321.1| (AE000222) orf, hypothetical
 protein [Escherichia coli]"
 /codon_start=1
 /transl_table=11
 /protein_id="AA059490.1"
 /db_xref="GI:12519315"
 /translation="MGAPVEMLEFRQTGTGARYLKDMDHGLQLFAIDAGQFRPPD
 EPELREYGSANTSTERSAYPRVRLVALMNLGITFY"
 complement(1271..2251)
 /gene="y1h8"
 /note="25905"
 complement(1271..2251)
 /gene="y1h8"
 /function="orf; unknown function"
 /note="Residues 1 to 326 of 326 are 97.54 pct identical to
 residues 1 to 326 from Escherichia coli K-12 strain
 MG1655: BA309"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC59491.1"
 /db_xref="GI:12519316"
 /translation="MNAITSPDYVYITVAGSNMAYGEGLELPREDAPHPRIKQL
 ARAHHPGSPSCHFNDIPIETPCHPDVDMOSTYHHPLATNHOTGYTGQALHARK
 LRPFPNACILIVPCCRGSATPTASSEGYSRHSASHDACRMGTDTPLDYDLSRT
 RAALVKRPQKFLGVCWMOGEFLDMTSDVSHQHNHNVAFRDLKQYHSLNNIT
 DAWPCDDTWTWKENPFAVEALVGNQNNILANTIFVDFQOQAGRLGNAPDEBDP
 DLSTGYGSAVRSPEWNTTAXRSRSHSSAARGLISDRVEAILQWRNR"
 complement(2316..3530)
 /gene="y1hT"
 /note="25906"
 complement(2316..3530)
 /gene="y1hT"
 /function="orf; unknown function"
 /note="Residues 1 to 404 of 404 are 98.26 pct identical to
 residues 1 to 404 of 404 from Escherichia coli K-12 strain
 MG1655: B4310"
 /codon_start=1
 /transl_table=11

CDS

gene

CDS

gene

CDS

gene

CDS

gene

misc-feature

FEATURES

source

TITLE

JOURNAL

SUBMISSION

WISCONSIN

```
/product="orf, hypothetical protein"
/protein_id="AAG59492.1"
/db_xref="GI:12519317"
/translation="MTTTRVRFPTTAETIIRKTVIALNCHLNCSSROETVMNKTITAL
ALMAEFANASVLPETVPFSGGAIIDNDVYIGLSAGTAWKLETOAKDKMTA
LAAPGPRDOATSAFIDNLIYFGGIGKNSGLTQVENDVHRIKNTKISVYKLSHA
PMGMAGHTFVHNGKAYTVGVGNONIFNGYFEDLNKAGDSTAVDK INAHYDKRAED
YFENKFLSEDSPTQOOSYAGESPMWGTAGAAVNVNGDKTWL INEAKPGTLDAVE
LPTGNNLKWNRLAPSSPDGYAGFAGISNDLIFAGCAGFGKSGRENOQNKYNAHE
GLKKSSTDIHLHMHNGKMDKSGELSGRAGVSLPMWNLSLLIGETAGCAKAVTDSL
ISVKNKNTVON
complement(3442..4167)
/gene="yjhA"
/note="25907"
complement(3442..4167)
/gene="yjhA"
/function="orf, Unknown function"
/note="Residues 1 to 241 of 241 are 100.00 pct identical
to residues 1 to 241 of 241 from Escherichia coli K-12
Strain MG1655: B4311"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG59493.1"
/db_xref="GI:12519318"
/translation="MKFMKRAKILSGVLLCFSSPLISOATIDVKGGRSGSHAYET
RLKVEGMOGMMWASMESNTWNTIHDKKENALINDVQEVNVAIKLDQWTPGMVL
THFSSNGTRGYPYKLSMDATKDNFGIRYRPMKAYROQDLSGDMSRDVHMDVY
TYHNSDFEAMOTITLTKSDNYRANHKWATENAFVLQYHMTPTITPEYDYDR
OGVYNGRDNLSNSYRIGVSFKL"
5614..6216
/gene="fimB"
/note="25910"
5614..6216
/gene="fimB"
/function="regulator; Cell exterior constituents: Surface
structures"
/note="Residues 1 to 200 of 200 are 100.00 pct identical
to residues 1 to 200 of 200 from Escherichia coli K-12
Strain MG1655: B4312"
/codon_start=1
/transl_table=11
/product="recombinase involved in phase variation;
regulator for fimA"
/protein_id="AAG59494.1"
/db_xref="GI:12519319"
/translation="MKNKADNKKRNFLTHSEISLKAANTGPHANNYCTLLCTFH
GFASEICRLISDIDLKACIYIHRLLKGFSTHPLINKVQALKWLSRTSYPA
ESEWFLSRKGNPLSROGFYHISTSGNAGLSLEIHPHMLRHS CGFALNMGIDTRL
IODYLGHNRHRTWYATASNAGREFYGLMDRARGORHNAVLT"
6694..7290
/gene="fime"
/note="25911"
6694..7290
/gene="fime"
/function="regulator; Cell exterior constituents: Surface
structures"
/note="Residues 1 to 198 of 198 are 99.49 pct identical to
residues 1 to 198 of 198 from Escherichia coli K-12 strain
MG1655: B4313"
/codon_start=1
/transl_table=11
/product="recombinase involved in phase variation;
regulator for fimA"
/protein_id="AAG59495.1"
/db_xref="GI:12519320"
/translation="MSKRRIYLTGCEVQAMQAVCGATGARDYCIILAYRHMRISE
LIDLHYODDLNEGRINIRLKNFSTVHPLIFEDREAVERTLERAWKADRTDAI
ETSRGSRSLRQOAVRITRDAGIEAGTVOTHPHMLRHACGYELAEBCADRLLIDYLL
GHRNIRHYRTYATASNAARFAGLWERNNILNKLKREEV"
7753..8303
/gene="fima"
/note="25912"
```

```
CDS
7755..8303
/gene="fima"
/function="structural component; Cell exterior
constituents: Surface structures"
/note="Residues 1 to 182 of 182 are 96.70 pct identical to
residues 1 to 182 of 182 from Escherichia coli K-12 Strain
MG1655: B4314"
/codon_start=1
/transl_table=11
/product="major type 1 subunit fimbria (pilin)"
/protein_id="AAG59496.1"
/db_xref="GI:12519321"
/translation="MKIKTALIVLSALSLSTAALAAATVNGCTVHFGEVYNAAC
AYDAGSVDTVOVLGOVPRTASLAODGATSSAVGPNIOIOLDCDTNASKAAVFLGTAID
ACHRTNVIALOSSAGSATNVGVOLLDRFGALLLDGATFSSQTLNNGTNIIPQARY
YALGEATPGANADATFVQYQ"
8260..8907
/gene="fimt"
/note="25913"
8260..8907
/gene="fimt"
/function="structural component; Cell exterior
constituents: Surface structures"
/note="Residues 1 to 215 of 215 are 98.60 pct identical to
residues 1 to 215 of 215 from Escherichia coli K-12 strain
MG1655: B4315"
/codon_start=1
/transl_table=11
/product="fimbrial protein"
```

```
Alignment Scores:
Pred. No.: 2, 67e-74 Length: 14466
Score: 824.00 Matches: 156
Percent Similarity: 97.52% Conservative: 1
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 97.17% Indels: 0
DB: Gaps: 1
```

US-09-900-575-29_Copy_26_186 (1-161) x AE005662 (1-14466).

```
OY 1 ProValValasnaValGlyGlnanleuValValAspenserhrgInlephesyHis 20
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13586 CCTCCCGTGAATGCGGCAAAACCTGCTGATGATTTTCGACGCAATCTTTGCCAT 13645
OY 21 AsnAspYrProGluTrnIleThrAspYrValThrleuGlnArgGlySerAlaTyrgly 40
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13646 AAGCATTAATCGGAACCATTAACAGTATGTCACACTGCACAGAGCGCTGCTATGCG 13705
OY 41 GlyValleuSerAsnPheserGlyThrValIlySerGlySerSerTyrrProPhePro 60
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13706 GCGGCTTATCTAATTTTCCGGGACCGTAATAATATAGTGGCAGTACATCCATTCCG 13765
OY 61 ThrThrSerGluTrnProArgValValTyrrAsnSerArgThrAspYrProTrrProVal 80
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13766 ACCACACGGAACGCGCGGCTGTTATTAATTCGGAACGATFAGCCGTCGCGCGTG 13825
OY 81 AlaIeuTyrrleuTrnProValSerSerAlaGlyIleuValIleIlyAlaGlySerleu 100
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13826 GCGCTTATTTGACGCGCTGAGCAGTGGCGGGGGGCGGATTAAGCTGGCTCATTTA 13885
OY 101 IleAlaValleuIleLeuArgGlnThrAsnAsnTyrrAsnSerAspPheGlnPheVal 120
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13886 ATTGCGGCTTATTTTTCGACAGACCAAAACATATAACGCAATGTTCCAGTTTGTG 13945
OY 121 TrpAsnIleTyrrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13946 TGGATATTTTACCAATAAAGATGTGTAGTGCCTACGCGGCTGGAGTTCCTGCT 14005
OY 141 ArgAspValThrValThrleuPrrAspYrArgGlySerValProIleProleuThrVal 160
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14006 CGTGATGCACCGCTACTCTGCCGAGCATCCCTGCTCAGTCCATTCCTTACCGCT 14065
OY 161 Tyr 161
```



```

/product="ornithine carbamoyltransferase 1"
/protein_id="BAB38654.1"
/db_xref="GI:13364709"
/translation="MSGFYHKKHFLKLLDTPPAEINSLLOLAKIKAKKSKGKEAKLT
GKNAIIFEDKSTRTKCSFEVAAYDQAKRTYILPGSGSITGKESIKPTAPYLGRYD
GIQYRGQGEIVETLAEYAGVPWNGLTVEFHPTQLADLTQEHLPKRPENRMTLY
YAGDARNMGNSMLEAALTLGLDLRLVAPQAPPEALVTECHALAQONGMNTLED
VAKGVGADIEYTDVWVSMGKAKEMWERTALRLRDYQVNSKMQLTQNGEYFLHCLP
AFPHDDQTTLGKNAEEFGHGMETDEVESAASTVFDQAEHRMHTIKAVVATLSLK
"
/gene="ECS5232"
/4284..4700
/gene="ECS5232"
/4284..4700
/note="similar to YUGD_ECOLI g11790704 percent identity
100 in 138 aa (conserved in E.coli K-12)"
/codon_start=1
/evidence="not experimental"
/translation="hypothetical protein"
/protein_id="BAB38655.1"
/db_xref="GI:13364710"
/translation="MANPQLEQREERTLIEELLEDSDPALYTIENHLSADDE
TLKAAVAERKLTGYTDPELEVEDGDIVICDDIISICALNMDLDAQVEQMLTAE
KFDVEYDGMGTFFEDPNEGDEDDFVDEDDDGVRH"
/complement(4746..5039)
/gene="ECS5233"
/complement(4746..5039)
/note="similar to BA256_ECOLI g11790705 percent identity
93 in 97 aa (conserved in E.coli K-12)"
/codon_start=1
/evidence="not experimental"
/translation="hypothetical protein"
/protein_id="BAB38656.1"
/db_xref="GI:13364711"
/translation="MWGGGTAIPLAGSFDICLOKMYFLPAIRGKGLAKKLAKAME
EAREGFRCYLETFLAKAIGLYEHLGFQHDIDYALGCTGHVDCVAMLEL"
/4442..6638
/gene="ECS5234"
/5442..6638
/note="similar to part (5-393 in 393 aa) of YUGN_ECOLI
g11267367 percent identity 96 in 389 aa (Conserved in
E.coli K-12)"
/codon_start=1
/evidence="not experimental"
/translation="hypothetical protein"
/protein_id="BAB38657.1"
/db_xref="GI:13364712"
/translation="WAQVYNEDVPSSHVSFVHGTERFELICVNVVLLTTITGITYLP
WALMKCKRYLANMEVNGRFSYGITGWFSECLVFFETAILMTYSADMPRLGCV
LTLSTLILFPAARGLRYOALMTSLNGVRFSCNMGMVWVTFDLPLAIGKGT
FPISTRMLHANSSSSVIVSVAIYVIGTISGTFSLVMSLWNTSGIHREF
VTLDPAYCIKVAIILAFIALPLAVAGYIIPQIINADSSVYADDIENLOQFMEMO
RKMTIROLIYRGIAVSTSYLVSIRNHFMNSLINDGIRFRSTLYTHGMLYRCAL
VVISGITGELATPLKIMIMTDQAKNTYILGDLDDLPLINKBEGPDKGLASISGIM

```

Alignment Scores:

```

Pred. No.: 4..52e-73 Length: 168764
Score: 824.00 Matches: 156
Percent Similarity: 97.52% Conservative: 1
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 97.17% Indels: 0
DB: 1 Gaps: 0

```

US-09-900-575-29_COPY_26_186 (1-161) x AP002569 (1-168764)

QY 1 ProValValAsnValIGlyGlnAsnLeuValAspLeuSerPheGlnIlePheCysHis 20
 DB 74151 CCTGCCGTGAATGGGGCAAAACCTGCTGATGCTTTGACGCAAAATCTTTGGCAT 74210

```

QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 74211 AACGATTATCCGGAACCAATTACAGACTATGTACACTGCACAGCGCTGCTTATGGC 74270
QY 41 GValLeuSerAsnPheSerGlyThrValIleTyrSerGlySerSerTyrProPhePro 60
DB 74271 GGCCTGTATCTAATTTTCCGGACCGTAATAATATAGGCAGTATCCATTCATTCGG 74330
QY 61 ThrThrSerGluThrProArgValIleTyrAsnSerArgThrAspLysProThrProVal 80
DB 74331 ACCACAGCGAAACGCCGCGGCTGTTATTAATTCGAGAACGATTAAGCGCGCGTG 74390
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
DB 74391 GCGCTTATTATGACGCTGTGACACTGCGGCGGGGTGGCATTAAAGCTGGCTCATTA 74450
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
DB 74451 ATTGCGCTGCTTATTTTGGCAGACCAAAACTATACAGCGCATGATTTCCAGTTTGTG 74510
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 74511 TCGAATATTACCGCAATTAATGATGTGTAGTGCCTACTGCGCGCTGCCATGTTCTGCT 74570
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 74571 CGTATGTACCGCTTACTGCGCGACTACCGTTCAGTCCAAATTCCTTACCGTT 74630
QY 161 Tyr 161
DB 74631 TAT 74633
RESULT 27
AF154927 1040 bp DNA linear BCT 10-JUL-2000
AF154927 Escherichia coli f1mH precursor (f1mH) gene, f1mH-244 allele,
DEFINITION complete cds.
ACCESSION AF154927
VERSION AF154927.1 GI:5524631
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1040)
AUTHORS Hamrick,T.S., Harris,S.L., Spears,P.A., Havelle,E.A., Horton,J.R.,
Russell,P.W. and Orndorff,P.E.
Genetic characterization of Escherichia coli type 1 pilus adhesin
mutants and identification of a novel binding phenotype
JOURNAL J. Bacteriol. 182 (14), 4012-4021 (2000)
MEDLINE 20327582
PUBMED 10869080
REFERENCE 2 (bases 1 to 1040)
AUTHORS Hamrick,T.S., Harris,S.L., Spears,P.A., Havelle,E.A., Horton,J.R.,
Russell,P.W. and Orndorff,P.E.
Direct Submission
Submitted (28-MAY-1999) Microbiology, Pathology, and Parasitology,
North Carolina State University, College of Veterinary Medicine,
4700 Hillsborough Street, Raleigh, NC 27606, USA
FEATURES
source location/Qualifiers
1..1040
/organism="Escherichia coli"
/strain="K12"
/db_xref="taxon:562"
47..949
/gene="f1mH"
/note="mutant allele"
/allele="f1mH-244"
47..949
/gene="f1mH"
/note="type 1 pil adhesin"
/codon_start=1

```

```
/transl_table=11
/product="FimH precursor"
/protein_id="AAD44321.1"
/db_xref="GI:5524632"
/translation="MKRVITLFAVLKMGVNMSPKCTANGTAIPIGGSANVYN
LAPVNVGNLVVDLSTQIFYHNDYPTETIDYVTLDRGSAYGVLSFGTSYSGSS
LPEPTSETPRVYNSRTDKPMFVALYLPVSSAGVALIAGSLIALILROTNVNS
DFOFWNNIYANNVYVPTGCDVSADYVTLPDYGSVPIPLTYVCAKSONLGYLL
SGTTADAGNSIFNTASFSPAGCVQVLTNGITIFANNVSLGAVGTSVSLGLTAN
YARRGGVTAAGNVSTIGVFVYQ"
240
variation
/gene="fimH"
/note="confers a cysteine to tyrosine change"
/replace="g"
BASE COUNT      253 a      244 c      269 g      274 t
ORIGIN
Alignment Scores:
Pred. No.:      1,64e-75      Length:      1040
Score:          823.00      Matches:      157
Percent Similarity: 98.14%      Conservative: 1
Best Local Similarity: 97.52%      Mismatches: 3
Query Match:    97.05%      Indels:      0
DB:             1      Gaps:      0
US-09-900-575-29_COPY_26_186 (1-161) x AF154927 (1-1040)
QY      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
Db      185 CCGCGTGTGAATGTGGGGCAAAACCTGTGTGATCTTTGACGCAATCTTTTACCAT 244
QY      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
Db      245 AACGATATTCGGGAACCATTAACACTATGTCACATGCAACGAGCGTGGCTTATGGC 304
QY      41 GlyValLeuSerAspPheSerGlyThrValLysTyrSerGlySerTyrProPhePro 60
Db      305 GCGGTGTATCTAATTTTTCGGGACCGTAATAATATAGTGGCAGTACATTCATTCC 364
QY      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrrProVal 80
Db      365 ACCACACGCAAAACGCGCGCGTGTATTAATTCGAGAACGATTAACCGTGGCGG 424
QY      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
Db      425 GCCCTTATTTGACGCGCTGTGACAGTGCAGCGCGGCGGCGATTAAAGCTGCTCATTA 484
QY      101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
Db      485 ATTGCGCTGTCTTATTTTGGCAGACCAACACTATTAACAGCGATATTCCAGTTTGTG 544
QY      121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
Db      545 TGGAAATATTACGCCAATATATGATGTGTGCTTACAGCGCGGCGCATGTTTCTGCT 604
QY      141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
Db      605 CGTGAATGTACCGTACTGTGCGGACTACCTGTTCAGTGCATTCCTTACCGTT 664
QY      161 Tyr 161
Db      665 TAT 667
RESULT 28
AX354050      837 bp      DNA      linear      PAT 06-FEB-2002
LOCUS      AX354050
DEFINITION      Sequence 16 from Patent WO0204496.
ACCESSION      AX354050
VERSION      AX354050.1 GI:18618987
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
```

```
REFERENCE      1 Escherichia.
AUTHORS      Langermann,S., Revel,A., Auguste,C. and Burllein,J.
TITLE      FimH adhesin proteins and methods of use
JOURNAL      Patent: WO 0204496-A 16 17-JAN-2002;
MEDLINE      MEDIMUNE, INC. (US)
FEATURES
Source      Location/Qualifiers
1..837
/organism="Escherichia coli"
/db_xref="taxon:562"
BASE COUNT      195 a      202 c      221 g      219 t
ORIGIN
Alignment Scores:
Pred. No.:      1.62e-75      Length:      837
Score:          822.00      Matches:      155
Percent Similarity: 98.14%      Conservative: 3
Best Local Similarity: 96.27%      Mismatches: 3
Query Match:    96.93%      Indels:      0
DB:             6      Gaps:      0
US-09-900-575-29_COPY_26_186 (1-161) x AX354050 (1-837)
QY      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
Db      76 CCGCGGTGAATGTGGGGCAAAACCTGTGTGATCTTTGACGCAATCTTTTGCCT 135
QY      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
Db      136 AACGATATTCGGGAACCATTAACACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195
QY      41 GlyValLeuSerAspPheSerGlyThrValLysTyrSerGlySerTyrProPhePro 60
Db      196 GCGGTGTATCTAATTTTTCGGGACCGTAATAATATAGTGGCAGTACATTCATTCC 255
QY      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrrProVal 80
Db      256 ACCACACGCAAAACGCGCGCGTGTATTAATTCGAGAACGATTAACCGTGGCGG 315
QY      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
Db      316 GCGCTTATTTGACGCGCTGTGACAGTGCAGTGGCGGCGGCGATTAAAGCTGCTCATTA 375
QY      101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
Db      376 ATGCGTGTCTAATTTTGGCAGACCAACTATTAACAGCGATATTCCAGTTTGTG 435
QY      121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
Db      436 TGGAAATATTACGCCAATATATGATGTGTGCTTACAGCGCGGCGCTGTGATTTCTCT 495
QY      141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
Db      496 CGTGAATGTACCGTACTGTGCGGACTACCTGTTCAGTGCATTCCTTACCGTT 555
QY      161 Tyr 161
Db      556 TAT 558
RESULT 29
AX354052      837 bp      DNA      linear      PAT 06-FEB-2002
LOCUS      AX354052
DEFINITION      Sequence 18 from Patent WO0204496.
ACCESSION      AX354052
VERSION      AX354052.1 GI:18618989
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
REFERENCE      1
AUTHORS      Langermann,S., Revel,A., Auguste,C. and Burllein,J.
TITLE      FimH adhesin proteins and methods of use
```

JOURNAL		Patient: MO 0204496-A 18 17-JAN-2002:	
FEATURES		MEDIMMUNE, INC. (US)	
source	Location/Qualifiers		
	1..837 /organism="Escherichia coli" /db_xref="taxon:562"		
BASE COUNT	188 a 202 c 226 g 221 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1.62e-75	Length:	837
Score:	822.00	Matches:	155
Percent Similarity:	98.14%	Conservative:	3
Best local Similarity:	96.27%	Mismatches:	3
Query Match:	96.93%	Indels:	0
DB:	6	Gaps:	0
US-09-900-575-29_COPY_26_186 (1-161) x AX354052 (1-837)			
Qy	1 ProValValAsnValIGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20		
Db	76 CCGTGGCGGAAATGTTGGGGCAAAACCTGGTCGCGATCTTTCAGACCAATCTTTGGCAT 135		
Qy	21 AsnAspYrProGluThrIleThrAspYrValIThrLeuGlnArgGlySerIleAtyGly 40		
Db	136 AACGGATTACCCGGAAACATTACAGACTATGTCCACACTGCACACGAGGTTGGCTATGGC 195		
Qy	41 GIValIleSerAsnPheserGlyThrValIleYrSerGlySerSerYrProPhePro 60		
Db	196 GCGCGTATATCAATGTTTTCGGGACCGGAAATATATATGCGCAGTATCCTTCCCT 255		
Qy	61 ThrThrSerGluThrProArgValValYrAsnSerArgThrAspLysProThrProVal 80		
Db	236 ACTACACAGGAACCCCGGGGCTTTTATAATTCAGAGAACGGATTAAGCCGTGGCGGTG 315		
Qy	81 AlaLeuYrLeuThrProValSerSerAlaGlyGlyLeuValIleYsAlaGlySerLeu 100		
Db	316 GCGCTTATTTACCGCTCTGTGACAGTGGGGGGGAGTGGCATTTAAACCTGGCTCATTA 375		
Qy	101 IleAlaValLeuIleLeuArgGlnThrAsnAsnYrAsnSerAspAspPheGlnPheVal 120		
Db	376 ATTTGGCGTCTTATTTTGGGACAGACCAACCTATTAACAGGAGATTTTCCAGTTTGG 435		
Qy	121 TrpAsnIleTyrAlaAsnAsnAspValValAlaProThrGlyGlyCysAspValSerAla 140		
Db	436 TGGAAATATTACGCCAATATATGATGTGGTGGGCCACACTGGCGCTGCATGTTTGCT 495		
Qy	141 ArgAspValThrValThrLeuProAspYrArgGlySerValProIleProLeuThrVal 160		
Db	496 CGTGATGTACACGTTACTCTGCCGCGACTCCCTGGTTCAGTCCGATTCCTCTTACCGTT 555		
Qy	161 Tyr 161		
Db	556 TAT 558		
RESULT 30 AF288194 937 bp DNA linear BCT 28-MAR-2001			
LOCUS AF288194			
DEFINITION Escherichia coli type 1 fimbrial adhesin subunit F1mH (f1mH) gene,			
complete cds.			
ACCESSION AF288194			
VERSION AF288194.1 GI:10946256			
KEYWORDS			
SOURCE			
ORGANISM			
Escherichia coli.			
Escherichia coli.			
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
Escherichia.			
1 (bases 1 to 937)			
REFERENCE			
AUTHORS Boudreau,J., Barnich,N. and Darfeuille-Michaud,A.			
TITLE Type 1 pili-mediated adherence of Escherichia coli strain LP82			
isolated from Crohn's disease is involved in bacterial invasion of			
intestinal epithelial cells			
JOURNAL Mol.Microbiol. 39 (5), 1272-1284 (2001)			

FEATURES	source
gene	<p>1. .937</p> <p>/organism="Escherichia coli"</p> <p>/strain="LF82"</p> <p>/db_xref="taxon:562"</p> <p>/note="Isolated from the ileal mucosa of a Crohn's disease patient"</p> <p>/gene="fimH"</p> <p>35. .937</p> <p>/gene="fimH"</p> <p>35. .937</p> <p>/gene="fimH"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="type 1 fimbrial adhesin subunit fimH"</p> <p>/protein_id="AAC24827.1"</p> <p>/db_xref="GI:10946257"</p> <p>/translation="MKRVLTFLFVLLMGWSVNAWSFACKTANGTAIPIDGGGSANVYVNLAAVAVGNLWDLSTQIFCHANDPEPTIDYVTLQSGAVGVLSFSGKYNKSSYTPPTTSETPRVYVNSTDKPMKPAIYLFVSSAGVATKAGSLIIVLLIROTNNYMSIDDFQVWNIYIANNVYVPTGGCCVSAADYVTLFDYGSVPIDLTIVCASONIGYLLSGQTADGNSIFNTTASFSPAGGVOLTRNGTIIIPANNVYSLGAVGSAVSLGLTANVARTGGVTAAGNVOSIIIGTVETVVO"</p>
CDS	<p>219 c 249 g 252 t</p>
BASE COUNT	217 a 219 c 249 g 252 t
ORIGIN	
Alignment Scores:	
Pred. No.:	1.84e-75
Score:	822.00
Percent Similarity:	98.14%
Best local Similarity:	96.27%
Query Match:	96.93%
DB:	1 Gaps: 0
US-09-900-575-29_COPY_26_186 (1-161) x AP288194 (1-937)	
QY	1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
Db	173 CCGTGGCGTGAATGTCGGGCAAAACCTGGTCGTGGATCTTTCGACGCCAAATCTTTGGCAT 232
QY	21 AsnAspTyrProGlnIleThrIleThrAspTyrValIlePheGlnArgGlySerAlaTyrGly 40
Db	233 AACGATVACCCGGAAACCACTTACAGACTATGTCACACTCCAAACGAGGTTCCGCTATGGC 292
QY	41 GlyValLeuSerAsnPheSerSerGlyIleThrValTyrSerGlySerSerTyrProPhePro 60
Db	293 GGGCGTATATCACTGTTTTCGGGACCGTAAATATTAATGACATGATATCTTTCCT 352
QY	61 ThrThrSerGlnThrProArgValValTyrAsnSerArgThrAspLysProIleProVal 80
Db	353 ACTACCAGGAAACCCCGGGGCTTTTATTAATTCGAGACGATTAAGCCGTGGCGGGTG 412
QY	81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
Db	413 GCGCTTATTTACCGCTGTGACGACTGGGGGGGAGTGGCGATTTAAACCTGGCTCATTA 472
QY	101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
Db	473 ATTGGCGTCTATTTTGGACAGCAACAACCTAATACGACGATCATTTCCAGTTGTG 532
QY	121 TrpAsnIleTyrAlaAsnAsnAspValValIleProThrGlyGlyCysAspValSerAla 140
Db	533 TGGAAATATTACGCCAATATATGATGGGGTGCCACACTGGCGCTGATGTCTTGCT 592
QY	141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160

Db 593 CGTATGTCACCGTACTTGGCCGACATACCTGGTTCAGTCCGATTCTTACCGTT 652
QY 161 Tyr 161
Db 653 TAT 655
RESULT 31
AX354042
LOCUS AX354042 837 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 8 from Patent WO0204496.
ACCESSION AX354042
VERSION AX354042.1 GI:18618979
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1
AUTHORS Langermann,S., Revel,A., Auguste,C. and Burtlein,J.
TITLE Film adhesin proteins and methods of use
JOURNAL Patent: WO 0204496-A 8 17-JAN-2002;
MEDIMUNE, INC. (US)
FEATURES
source Location/Qualifiers
1..837
BASE COUNT 192 a 200 c 225 g 220 t
ORIGIN /organism="Escherichia coli"
/db_xref="taxon:562"
Alignment Scores:
Pred. No.: 2.6e-75 Length: 837
Score: 820.00 Matches: 155
Percent Similarity: 96.89% Conservative: 1
Best Local Similarity: 96.27% Mismatches: 5
Query Match: 96.70% Indels: 0
DB: Gaps: 6
US-09-900-575-29_COPY_26_186 (1-161) x AX354042 (1-837)
QY 1 ProValValAsnValGlyGlnAsnLeuValAlaSpluSerThrGlnIlePheCysHis 20
Db 76 CCCCCCGTGAATGGGGGCAAAACCTGGTCGATCTTTCGACGCAAACTTTGCCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
Db 136 AAGCATATCCGGAACCATTAACAGACTATGTCACACTGCACGAGGCTCGGCTTATGGC 195
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
Db 196 GGGGTGTATCTAATTTTCCGGGACCGTAAATATAGTGGCAGTAGCTATCCATTTCCG 255
QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrrProVal 80
Db 256 ACTACCAAGCAAAACGCGGGGTTGTTTAAATTCGAGAACGGAATMAAGCCGTGGCGGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLysLeuValIleLysAlaGlySerLeu 100
Db 316 GCGCTTATTTGACGCGCTGAGCACTGGGGTGGGCGCATTAACCTGGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
Db 376 ATTCGCGTCTTATTTTGGGACAGACCAACTATTAACAGCAGATTTTCCAGTTTGG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValIleProThrGlyGlyCysAspValSerAla 140
Db 436 TGGAAATATTAGCCCAATATATGATGGTGGCTTACTGGGCGCTGCAGTGTCTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
Db 496 CATGATGTACCGTACTCTGCCGAGACTACCTGGTTCAGTCCCAATTCTTACCGTT 555
QY 161 Tyr 161
|||

Db 556 TAT 558
RESULT 32
AX354045
LOCUS AX354045 837 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 11 from Patent WO0204496.
ACCESSION AX354045
VERSION AX354045.1 GI:18618982
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1
AUTHORS Langermann,S., Revel,A., Auguste,C. and Burtlein,J.
TITLE Film adhesin proteins and methods of use
JOURNAL Patent: WO 0204496-A 11 17-JAN-2002;
MEDIMUNE, INC. (US)
FEATURES
source Location/Qualifiers
1..837
BASE COUNT 193 a 202 c 224 g 218 t
ORIGIN /organism="Escherichia coli"
/db_xref="taxon:562"
Alignment Scores:
Pred. No.: 3.29e-75 Length: 837
Score: 819.00 Matches: 154
Percent Similarity: 98.14% Conservative: 4
Best Local Similarity: 95.65% Mismatches: 3
Query Match: 96.58% Indels: 0
DB: Gaps: 6
US-09-900-575-29_COPY_26_186 (1-161) x AX354045 (1-837)
QY 1 ProValValAsnValGlyGlnAsnLeuValAlaSpluSerThrGlnIlePheCysHis 20
Db 76 CCTGCCGTGAATGGGGCAAAACCTGGTCGATCTTTCGACGCAAACTTTGCCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
Db 136 AAGCATATCCGGAACCATTAACAGACTATGTCACACTGCACGAGGCTCGGCTTATGGC 195
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
Db 196 GGGGTGTATCTAATTTTCCGGGACCGTAAATATAGTGGCAGTAGCTATCCATTTCC 255
QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrrProVal 80
Db 256 ACTACCAAGCAAAACGCGGGGTTGTTTAAATTCGAGAACGGAATMAAGCCGTGGCGGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLysLeuValIleLysAlaGlySerLeu 100
Db 316 GCGCTTATTTGACGCGCTGAGCACTGGGGTGGGCGCATTAACCTGGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
Db 376 ATTCGCGTCTTATTTTGGGACAGACCAACTATTAACAGCAGATTTTCCAGTTTGG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValIleProThrGlyGlyCysAspValSerAla 140
Db 436 TGGAAATATTAGCCCAATATATGATGGTGGCTTACTGGGCGCTGCAGTGTCTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
Db 496 CGTATGTACCGTACTCTGCCGAGACTACCTGGTTCAGTCCCAATTCTTACCGTT 555
QY 161 Tyr 161
|||
RESULT 33
AX354056

LOCUS AX354056 837 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 22 from Patent WO0204496.
ACCESSION AX354056
VERSION AX354056.1 GI:18618993
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE
1 Langermann, S., Revel, A., Auguste, C. and Burieln, J.
TITLE Film adhesin proteins and methods of use
JOURNAL Patent: WO 0204496-A 22 17-JAN-2002;
MEDIMUNE, INC. (US)
FEATURES
source Location/Qualifiers
1..837
BASE COUNT 193 a 202 c 224 g 218 t
ORIGIN

Alignment Scores:
Pred. No.: 3,29e-75 Length: 837
Score: 819.00 Matches: 154
Percent Similarity: 98.14% Conservative: 4
Best Local Similarity: 95.65% Mismatches: 3
Query Match: 96.58% Indels: 0
DB: Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AX354056 (1-837)

QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 76 CCTGCCGTGAATGTGGGCAAAACCTGTCGTAGATCTTTGCAGCAATCTTTGCCAT 135
QY 21 AsnAspTyrProGlnThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 136 AACGATTACCCAGAAACCATTTACAGACTATGTCACACTGCAGCAAGAGTGGCGCTATGGC 195
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
DB 196 GCGCTGTATCTAGTTTTCCTGGGACCCGTAAATATTAATGAGCATGATCTCTTCCCT 255
QY 61 ThrThrSerGlnThrProArgValValTyrAsnSerArgThrAspLysProThrProVal 80
DB 256 ACTACACGCGAAACGCCGGGGTGTATTATTCAGAGAAAGGATTAAGCCGTGGCCGGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
DB 316 GCGCTTATTGTGACGCCGTGAGCACTGCGGGGAGTGGCATTAAGCTGGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
DB 376 ATTGCCGTGTATTTTGGACAGACCAACACTATTAACAGCATTTCCAGTTTGTG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 436 TGGATATTACCCCAATATGATGTGGTGTCCCATGCGGGCTGCATGTCTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 496 CGTGATGTCAACCGTACTCTGCGGACTACCTGTGTCAAGTCCGATCTTACCGTT 555
QY 161 Tyr 161
DB 556 TAT 558

RESULT 34
LOCUS AX354038 840 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 4 from Patent WO0204496.
ACCESSION AX354038
VERSION AX354038.1 GI:18618975

KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE
1 Langermann, S., Revel, A., Auguste, C. and Burieln, J.
TITLE Film adhesin proteins and methods of use
JOURNAL Patent: WO 0204496-A 4 17-JAN-2002;
MEDIMUNE, INC. (US)
FEATURES
source Location/Qualifiers
1..840
BASE COUNT 195 a 202 c 224 g 219 t
ORIGIN

Alignment Scores:
Pred. No.: 3.3e-75 Length: 840
Score: 819.00 Matches: 154
Percent Similarity: 98.14% Conservative: 4
Best Local Similarity: 95.65% Mismatches: 3
Query Match: 96.58% Indels: 0
DB: Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AX354038 (1-840)

QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 76 CCTGCCGTGAATGTGGGCAAAACCTGTCGTAGATCTTTGCAGCAATCTTTGCCAT 135
QY 21 AsnAspTyrProGlnThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 136 AACGATTACCCAGAAACCATTTACAGACTATGTCACACTGCAGCAAGAGTGGCGCTATGGC 195
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
DB 196 GCGCTGTATCTAGTTTTCCTGGGACCCGTAAATATTAATGAGCATGATCTCTTCCCT 255
QY 61 ThrThrSerGlnThrProArgValValTyrAsnSerArgThrAspLysProThrProVal 80
DB 256 ACTACACGCGAAACGCCGGGGTGTATTATTCAGAGAAAGGATTAAGCCGTGGCCGGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
DB 316 GCGCTTATTGTGACGCCGTGAGCACTGCGGGGAGTGGCATTAAGCTGGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
DB 376 ATTGCCGTGTATTTTGGACAGACCAACACTATTAACAGCATTTTCCAGTTTGTG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 436 TGGATATTACCCCAATATGATGTGGTGTCCCATGCGGGCTGCATGTCTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 496 CGTGATGTCAACCGTACTCTGCGGACTACCTGTGTCAAGTCCGATCTTACCGTT 555
QY 161 Tyr 161
DB 556 TAT 558

RESULT 35
LOCUS AF089840 982 bp DNA linear BCT 16-SEP-1998
DEFINITION Escherichia coli filmh (filmh) gene, complete cds.
ACCESSION AF089840
VERSION AF089840.1 GI:3599570
KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

REFERENCE	1 (bases 1 to 982)
AUTHORS	Pouttu, R. and Korhonen, T. K.
TITLE	Direct Submission
JOURNAL	Submitted (02-SEP-1998) Department of Biosciences, Division of General Microbiology, University of Helsinki, Viikinkaari 9 C, P.O.Box 56, Helsinki FIN-00014, Finland
FEATURES	location/Qualifiers
SOURCE	

BASE COUNT	236	a	232	c	257	g	257	t
ORIGIN								

Alignment Scores:	
Pred. No.:	3 95e-75
Score:	819.00
Percent Similarity:	98.14%
Best Local Similarity:	95.65%
Query Match:	96.55%
DB:	1
Length:	982
Matches:	154
Conservative:	4
Mismatch:	3
Indels:	0
Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AF089840 (1-982)

OY	1	ProValAlaSerValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis	20
Db	185	CTGCGCTGAATGGGGCAAAACCTGGTCGTAGATCTTTGGACGCAAACTTTTCCAT	24
OY	21	AsnAspTyrProGluThrIleThrAspTyrValThrIleGlnAcGlySerAlaTyrGly	40
Db	245	AACGATTACCCAAACCACTTACACACTATGACACACTGCAMCGAGGTGGCGTTATGGC	30
OY	41	GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro	60
Db	305	GGCGGTGATTACTAGTTTTCGGGACCGGTAATAATATGGCACTACTCTTCCTTCCCT	36
OY	61	ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProThrProVal	80
Db	365	ACTACACAGCAAAACGCCGGGGTGTATTATATTCGAGAACGGATTAAGCCGTGGCGGTG	42
OY	81	AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu	10
Db	425	GCCCTTATTATTTGACGGCGGTGACGTGCGGGGGAGCTGCGATTAAAGCTGGCTCATTA	48
OY	101	IleAlaValLeuIleLeuArgGlnThrAsnAspTyrAsnSerAspAspPheGlnPheVal	12
Db	485	ATTGGCGTGCCTTATTTTGGACAGACGCACAACTATTAACGCCATATTTTCCAGTTTG	54
OY	121	TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla	14
Db	545	TGGAAATATTTACGCCAATAATATGATGTGGTGGCCCTGCGCGGTGCATGTTCTGCT	60
OY	141	ArgAspValThrValThrLeuProAspTyrArgGlySerAlaProIleProLeuThrVal	16
Db	605	CGGATGTCACCGTTACTCTCGCGACTACCCCTGATGGTTCACGGCCGATTCCTTACCGTT	66

Qy	161	Tyr	161
		!!!	
Db	665	TAT	667

LOCUS	AF317710	1008 bp	DNA	linear	BCT 31-JAN-2001
DEFINITION	<i>Escherichia coli</i> f1mH (f1mH) gene, complete cds.				
ACCESSION	AF317710				
VERSION	AF317710.1	GI:12620397			
KEYWORDS					
SOURCE	<i>Escherichia coli</i> .				
ORGANISM	<i>Escherichia coli</i>				

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
1	(bases 1 to 1008)			2	(bases 1 to 1008)		
	Elliot,S.J. and Kim,K.S.	Virulence factors of meningitic <i>Escherichia coli</i>			Elliot,S.J. and Kim,K.S.	Direct Submission	
	Unpublished				Submitted (31-OCT-2000)	Pediatric Infectious Disease, Johns Hopkins	

BASE COUNT	246 a	236 c	263 g	263 t
ORIGIN				

Alignment Scores:	
Pred. No.:	4,086-75
Score:	819.00
Percent Similarity:	98.14%
Best Local Similarity:	93.65%
Query Match:	96.58%
BB:	1
Gaps:	0
Matches:	1008
Mismatch:	154
Indels:	3
MisMatches:	4
Indels:	0
Gaps:	0

US-09-900-5/5-29_COPY_26_186 (1-161) x AF317710 (1-1008)

OY		ProValAsnValGlyGlnAsnLeuValAspLeuSerThrGlnIlePheCysHis	20
Db	191	CCCGCCGGATGTGGGGCAAAACCTGTGTGAATCTTTCACCAGCAATCTTTGCCAT	25
OY	21	AsnAspTryProGluThrIleThrAspTryValThrLeuGlnArgGlySerAlaIrrGly	40
Db	251	AACGATTACCCGAGAACCATTAACACTATGTACACTCAACGAGGGCGGTTAAGGC	310
OY	41	GlyValLeuSerAsnPheSerGlyThrValLysTyrrSerGlySerSerTryProPhePro	60
Db	311	GCGCGTTATCATGTTTTTCGGGACCGTAAATATAATGACGATGACTATTCCTTCCCT	370

Alignment Scores:

Pred. No.: 4,17e-75 Length: 837
 Score: 818.00 Matches: 154
 Percent Similarity: 97.52% Conservative: 3
 Best Local Similarity: 95.65% Mismatches: 4
 Query Match: 96.46% Indels: 0
 DB: 6 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AX354043 (1-837)

QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
 DB 76 CCCGCTGTAATGTTGGGCAAAACCTGCTGATCTTTGACGCAAAATCTTTGCCAT 135
 QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
 DB 136 AACGATTAATCCGGAACCAATTAACAGACTAATGTCACACTGACAGAGAGGTGCGCTTATGCG 195
 QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
 DB 196 GGCCTGTAATCTAATGTTTCCGGGACCGCTAAATATATATGACAGTATCTCTTCCCT 255
 QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTTPProVal 80
 DB 256 ACTACACGAGAAACGCGCGGCTTTATTAATTCGAGAACGAGATTAAGCCGCGCGGTG 315
 QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
 DB 316 GCGCTTATTTGACCGCGGTGAGACAGTGGGGGAGTGGCATTAAGCTGGCTCATTA 375
 QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
 DB 376 ATTGCCGCTGCTTAATTTGGCAGACCAACACTATTAACAGCATATTTCCAGTTTGTG 435
 QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
 DB 436 TGGAAATATTACCAATTAATGATGGTGGTCCCACTGGCGCTGATGCTTGTGCT 495
 QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
 DB 496 CGTGAATGTCACGCTTAATTTGGCGGACTACCTGCTTCACTCCGATCTTACCGTT 555
 QY 161 Tyr 161
 DB 556 TAT 558
 RESULT 39
 LOCUS AX354035 837 bp DNA linear PAT 06-FEB-2002
 DEFINITION Sequence 1 from Patent WO0204496.
 ACCESSION AX354035
 VERSION AX354035.1 GI:18618972
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE
 1 Langermann, S., Revel, A., Auguste, C. and Burieln, J.
 TITLE Film adhesin proteins and methods of use
 JOURNAL Patent: WO 0204496-A 1 17-JAN-2002;
 MEDIMUNE, INC. (US)
 FEATURES
 source 1..837
 Location/Qualifiers
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 BASE COUNT 186 a 204 c 228 g 219 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5.29e-75 Length: 837
 Score: 817.00 Matches: 155
 Percent Similarity: 96.89% Conservative: 1

Best Local Similarity: 96.27% Mismatches: 5
 Query Match: 96.34% Indels: 0
 DB: 6 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AX354035 (1-837)

QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
 DB 76 CCCGCTGTAATGTTGGGCAAAACCTGCTGATCTTTGACGCAAAATCTTTGCCAT 135
 QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
 DB 136 AACGATTAATCCGGAACCAATTAACAGACTAATGTCACACTGACAGAGGTGCGCTTATGCG 195
 QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
 DB 196 GGCCTGTAATCTAATGTTTCCGGGACCGCTAAATATATATGACAGTATCTCTTCCCT 255
 QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTTPProVal 80
 DB 256 ACCACGAGGAAACGCGCGGCTTTATTAATTCGAGAACGAGATTAAGCCGCGCGGTG 315
 QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
 DB 316 GCGCTTATTTGACCGCTTGTGACACTGGGGGAGTGGCATTAAGCTGGCTCATTA 375
 QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
 DB 376 ATTGCCGCTGCTTAATTTGGCAGACCAACACTATTAACAGCATATTTCCAGTTTGTG 435
 QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
 DB 436 TGGAAATATTACCAATTAATGATGGTGGTCCCACTGGCGCTGATGCTTGTGCT 495
 QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
 DB 496 CGTGAATGTCACGCTTAATTTGGCGGACTACCTGCTTCACTCCGATCTTACCGTT 555
 QY 161 Tyr 161
 DB 556 TAT 558
 RESULT 40
 LOCUS AX354047 837 bp DNA linear PAT 06-FEB-2002
 DEFINITION Sequence 13 from Patent WO0204496.
 ACCESSION AX354047
 VERSION AX354047.1 GI:18618984
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE
 1 Langermann, S., Revel, A., Auguste, C. and Burieln, J.
 TITLE Film adhesin proteins and methods of use
 JOURNAL Patent: WO 0204496-A 13 17-JAN-2002;
 MEDIMUNE, INC. (US)
 FEATURES
 source 1..837
 Location/Qualifiers
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 BASE COUNT 195 a 200 c 222 g 220 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6.7e-75 Length: 837
 Score: 816.00 Matches: 154
 Percent Similarity: 97.52% Conservative: 3
 Best Local Similarity: 95.65% Mismatches: 4
 Query Match: 96.23% Indels: 0
 DB: 6 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AX354047 (1-837)

QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
 DB 76 CCGGCCGTGATGGGGGCAAAACCTGCTGATCTTTCCAGCAACATCTTTGGCAT 135
 QY 21 AsnAspTyrProGlnThrIleThrAspTyrValIleuGlnArgGlySerAlaTyrGly 40
 DB 136 AACGATTACCCAGAAACCATTTACAGACTATGTCACACTGCACAGAGGCTTGATGGC 195
 QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
 DB 196 AGCGTGTATCTAGTTTTCGGGACCGCTAAATATATATGCACTACTCTTCCTTCCCT 255
 QY 61 ThrThrSerGlnThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
 DB 256 ACTACACGCAAGACGCGCGGTGTTTATATTCGAGAACGCGATTAAGCCGTCGGTG 315
 QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
 DB 316 GCGCTTATTTGACGCGCGGTGACGAGTGGGGAGTGGGATTAAAGCTGGCTCATTA 375
 QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
 DB 376 ATTGCGCGCTTATTTTGGCAGACACCAACTATTAACAGCATGATTCACGTTTGTG 435
 QY 121 TrpAsnIleTyrAlaAsnAsnAspValValIleProThrGlyGlyCysAspValSerAla 140
 DB 436 TGGATATTATTCGCCCAATATGATGTGTGTCGCCACTGGCGGCTGTGATGTTTGTCT 495
 QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
 DB 496 CGTGATGTACCGTTACTCTGCCGACTACCTGGTTTCAGTCCGATTCCTTACCGCTT 555
 QY 161 Tyr 161
 DB 556 TAT 558

RESULT 41
 LOCUS AX354048 837 bp DNA linear PAT 06-FEB-2002
 DEFINITION Sequence 14 from Patent WO204496.
 ACCESSION AX354048
 VERSION AX354048.1 GI:18618985
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli.
 Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE
 AUTHORS 1 Langermann, S., Revel, A., Auguste, C. and Burtlein, J.
 TITLE Film adhesin proteins and methods of use
 JOURNAL Patent: WO 0204496-A 14 17-JAN-2002;
 MEDIMUNE, INC. (US)

FEATURES
 source 1..837
 location/Qualifiers
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 BASE COUNT 195 a 200 c 222 g 220 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6.7e-75 Length: 837
 Score: 816.00 Matches: 154
 Percent Similarity: 97.52% Conservative: 3
 Best Local Similarity: 95.65% Mismatches: 4
 Query Match: 96.23% Indels: 0
 DB: 6 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x AX354048 (1-837)

QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
 DB 76 CCGGCCGTGATGGGGGCAAAACCTGCTGATCTTTCCAGCAACATCTTTGGCAT 135

DB 76 CCGGCCGTGATGGGGGCAAAACCTGCTGATCTTTCCAGCAACATCTTTGGCAT 135
 QY 21 AsnAspTyrProGlnThrIleThrAspTyrValIleuGlnArgGlySerAlaTyrGly 40
 DB 136 AACGATTACCCAGAAACCATTTACAGACTATGTCACACTGCACAGAGGCTTGATGGC 195
 QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
 DB 196 AGCGTGTATCTAGTTTTCGGGACCGCTAAATATATATGCACTACTCTTCCTTCCCT 255
 QY 61 ThrThrSerGlnThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
 DB 256 ACTACACGCAAGACGCGCGGTGTTTATATTCGAGAACGCGATTAAGCCGTCGGTG 315
 QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
 DB 316 GCGCTTATTTGACGCGCGGTGACGAGTGGGGAGTGGGATTAAAGCTGGCTCATTA 375
 QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
 DB 376 ATTGCGCGCTTATTTTGGCAGACACCAACTATTAACAGCATGATTCACGTTTGTG 435
 QY 121 TrpAsnIleTyrAlaAsnAsnAspValValIleProThrGlyGlyCysAspValSerAla 140
 DB 436 TGGATATTATTCGCCCAATATGATGTGTGTCGCCACTGGCGGCTGTGATGTTTGTCT 495
 QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
 DB 496 CGTGATGTACCGTTACTCTGCCGACTACCTGGTTTCAGTCCGATTCCTTACCGCTT 555
 QY 161 Tyr 161
 DB 556 TAT 558

RESULT 42
 LOCUS ECEFM 3356 bp DNA linear BCT 29-MAR-1999
 DEFINITION Escherichia coli fimD, fimF, fimG, fimH, uxax & gntP genes.
 ACCESSION AJ225176
 VERSION AJ225176.1 GI:3286745
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli.
 Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE
 AUTHORS 1 Klemm, P. and Christiansen, G.
 TITLE Three fim genes required for the regulation of length and mediation
 JOURNAL of adhesion of Escherichia coli type 1 fimbriae
 MEDLINE 88038337
 PUBMED 2890081
 REFERENCE 2 (bases 1 to 3356)
 AUTHORS Klemm, P., Tony, S., Nielsen, H. and Conway, T.
 TITLE The gntP gene of Escherichia coli involved in gluconate uptake
 JOURNAL J. Bacteriol. 178 (1), 61-67 (1996)
 MEDLINE 8550444
 PUBMED 8550444

REFERENCE 3 (bases 1 to 3356)
 AUTHORS Marc, D. and Dho-Moulin, M.
 TITLE Analysis of the fim cluster of an avian O2 strain of Escherichia
 coli: serogroup-specific sites within fimA and nucleotide sequence
 of fimI

JOURNAL J. Med. Microbiol. 44 (6), 444-452 (1996)
 MEDLINE 96235989
 PUBMED 8636962
 REFERENCE 4 (bases 1 to 3356)
 AUTHORS Blattner, F., Plunkett, G., Bloch, C., Perna, N., Burland, V., Riley, M.,
 Collado-Vides, J., Glasner, J., Rode, C., Mayhew, G., Gregor, J.,
 Davis, N., Kirkpatrick, H., Goeden, M., Rose, D., Mau, B. and Shao, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 Science 277 (5331), 1453-1474 (1997)

JOURNAL TITLE
 DB 76 CCGGCCGTGATGGGGGCAAAACCTGCTGATCTTTCCAGCAACATCTTTGGCAT 135

MEDLINE 97426617
PUBMED 9278503
REFERENCE 5 (bases 1 to 3356)
AUTHORS Marc.D., Arne.P., Bree.A. and Dho-Moulin.M.
TITLE Colonization ability and pathogenic properties of a fim- mutant of
JOURNAL an avian strain of Escherichia coli
MEDLINE 98439336
PUBMED 9766199
REFERENCE 6 (bases 1 to 3356)
AUTHORS Marc.D.
TITLE Direct Submission
JOURNAL Submitted (30-Dec-1997) Station de Pathologie Aviaire et
Parasitologie, Institut National de la Recherche Agronomique, INRA
- Centre de Tours, Nouzilly 37380, France
FEATURES
source 1. 3356
/organism="Escherichia coli"
/strain="MT78"
/db_xref="taxon:562"
/note="avian pathogenic strain of O2 serogroup (citation
[2])"
1. .264
/gene="fimD"
<1. .264
/gene="fimD"
/codon_start=1
/transl_table=1
/product="FimD protein"
/protein_id="CA12420.1"
/db_xref="GI:3286746"
/translation="ARVGKILMTLTTHNNKPLPFGAMVTSSOSSSGIIVADNGQVYLS
GMPAGKVVQWKEENAHCVANYQLPPESQOOLLQLSAECR"
274. .804
/gene="fimF"
274. .804
/gene="fimF"
/citation=[1]
/codon_start=1
/transl_table=1
/product="minor fimbrial subunit, precursor polypeptide"
/protein_id="CA12421.1"
/db_xref="GI:3286747"
/db_xref="SPTREMBL:O87737"
/translation="MRNKPFYLCAFLVLAASHALADSTTINGYVRDNGCSVAES
TNPYDLMENAKQENNICATTPVPEFRILLSPCGNAVSAYVGFQVADSHNALLA
LEHTVSAAGLGIQLNEQONEIPLNASSAISMTTLTPGKNTLNFYARLMTQVPY
TAGHNATATFTLEVO"
817. .1320
/gene="fimG"
817. .1320
/gene="fimG"
/citation=[1]
/codon_start=1
/transl_table=1
/product="minor fimbrial subunit, precursor polypeptide"
/protein_id="CA12422.1"
/db_xref="GI:3286748"
/db_xref="SPTREMBL:O87738"
/translation="MKWCKRGYLLAAMLAFAASATIQADVTITNGKVVAKPCTVST
NATVDLGLYSLSAGASAMHVDVALELNCVGTGSRVTSASPSGAADSGYKKKG
TAQNIQLELDSDGNTLNTGATKTVQVDDSSQSAHPLQVVALIVNGCATQGTQAVI
SITYYS"
1340. .2242
/gene="fimH"
1340. .2242
/gene="fimH"
/codon_start=1
/transl_table=1
/product="type 1 fimbriae adhesin, precursor polypeptide"
/protein_id="CA12423.1"
/db_xref="GI:3286749"
/translation="MKRYITLFAVLKMGVSNAMSFACKTANGTAIPIGGSANVYVN

LAPAVNNGOKLVYDLSSTOIECHNDYEPETITDVTYLQRCAGVUSSSEGTWKYNGSS
LPPPTSERPRVRYNRSRTDKPRPVALITLPVSSAGVAIKASLIATLRLQNTNYS
DDQCFWNITANDVVPVPGCGVSAADVTYVLPPGSPVPIPLTVYCAKSQLNGYL
SGTTADAGNSIFPTNASFSPAGGVQVLRNGITIIIPANTVSLAGVTSVSLGLTAN
YATGGQVTAAGVQSSIIIGTVFYQ"
2415. .2631
/gene="gntP"
2415. .2631
/gene="gntP"
/note="3' and 5' regions of gene gntP (most of the gene
deleted between nt 2523 and 2568)"
/pseudo
2971. .3356
/gene="uxaA"
2971. .3356
/gene="uxaA"
/citation=[4]
/codon_start=1
/transl_table=1
/product="uxaA protein"
/protein_id="CA12424.1"
/db_xref="GI:3286750"
/db_xref="SPTREMBL:O87739"
/translation="MEQTRWRYGPNDPVSLADVROAGATGVYTLAHHIPNCEVSWEE
ILKRAIVEDAGLWSVSEYPIHEDITHTHNGYEQWLTANYQOTLRNLACGIRTVYX
NEMPVLDWTRFDLEVLDPDGSKALRF"
BASE COUNT 845 a 819 c 866 g 826 t
ORIGIN
Alignment Scores:
Pred. No.: 3,31e-74 Length: 3356
Score: 816.00 Matches: 154
Percent Similarity: 97.52% Conservative: 3
Best Local Similarity: 95.65% Mismatches: 4
Query Match: 96.23% Indels: 0
DB: 1 Gaps: 0
US-09-900-575-29_COPY_26_186 (1-161) x ECFIM (1-3356)
QY 1 ProvalValasnaValGlyGlnasnaLeuValAspleuSerThrgInIlePhcGyaHis 20
DB 1478 CTTGCCGCGAATGTGGCGCAAAAGCTGCTGAGATCTTTCGACGCAATCTTTGCCAT 1357
QY 21 AsnaSPtyrProGluThrIleThrAspTyValThrIleuGlnarGlySerAlaTyGly 40
DB 1538 AACGATACCCAGAAACCATTAACAGACTATGTCACACTGCAACGAGGTCGCTTATGGC 1597
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTySerGlySerSertyrProPhero 60
DB 1598 GCGGTGTATCTAGTCTTTCGCGGACCGTAAATATATATGACGTATCTTTCCT 1657
QY 61 ThrThrSerGluThrProArGValValTyraSnaSerarGThrAspLysProTTPProVal 80
DB 1658 ACTACACACGAAACGCGCGGCTGTATTAATTCGAAACGATATAGCGGTCGCGCTG 1717
QY 81 AlaLeuTyrlleuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
DB 1718 GCGCTTATTTGACGCGCGGAGCAGTGCAGGGAATGCGATTAAAGCTGCTCATTTA 1777
QY 101 IleAlaValLeuIleLeuArGlnThrAsnaSntyrAsnaSerAspAspPheGlnPheVal 120
DB 1778 ATTCCGCTGCTTATTTTGCACAGACCAACATATTAACGCGATATCCAGTTTGTG 1837
QY 121 TrpAsnIleTyrlaAsnaSnaSnaSnaValValProThrglyGlycysAspValSerAla 140
DB 1838 TCGAATATTTACGCCCAT 1897
QY 141 ArgAspValThrValThrLeuProAspTyraGlySerValProIleProleuThrVal 160
DB 1898 CGTGATGTACCGTTACTGTGCGGACTACCGTACCGTACCGTACCGTACCGTACCGT 1957
QY 161 Tyr 161
|||

VERSION AF154929.1 GI:5524635
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1039)
AUTHORS Hamrick,T.S., Harris,S.L., Spears,P.A., Havelle,E.A., Horton,J.R.,
Russell,P.W. and Ondonoff,P.E.
TITLE Genetic characterization of Escherichia coli type 1 pilus adhesin
mutants and identification of a novel binding phenotype
JOURNAL J. Bacteriol. 182 (14), 4012-4021 (2000)
MEDLINE 20327582
PUBMED 10869080
REFERENCE 2 (bases 1 to 1039)
AUTHORS Hamrick,T.S., Harris,S.L., Spears,P.A., Havelle,E.A., Horton,J.R.,
Russell,P.W. and Ondonoff,P.E.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1999) Microbiology, pathology, and Parasitology,
North Carolina State University, College of Veterinary Medicine,
4700 Hillsborough Street, Raleigh, NC 27606, USA
location/Qualifiers
1. 1039
/organism="Escherichia coli"
/strain="K12"
/db_xref="taxon:562"
47. 436
/gene="flmH"
/note="mutant allele"
/allele="flmH-208"
47. 436
/gene="flmH"
/transl_start=1
/codon_start=1
/product="FlmH precursor"
/protein_id="AAD4323.1"
/db_xref="GI:5524636"
/translation="MKRVITLFAVLDMGMSVNAVSFACKTANGTAIPIGGSANVYV
LAPVNVGONLVLDSTQIFCHNDYDEITDVTYLTORGSAVGVSINFGSTVNIYAVA
IHPLPPAKRRALPIIRERISRGHWRFI"
337
variation
/gene="flmH"
/note="deletion creates in a frameshift: results in a
truncated protein"
BASE COUNT 251 a 244 c 270 g 274 t
ORIGIN
Alignment Scores:
Pred. No.: 7,26e-74 Length: 1039
Score: 807.00 Matches: 157
Percent Similarity: 98.14% Conservative: 1
Best Local Similarity: 97.52% Mismatches: 3
Query Match: 95.17% Indels: 1
DB: 1 Gaps: 0
US-09-900-575-29_COPY_26_186 (1-161) x AF154929 (1-1039)
QY 1 ProValValAsnValGlyGlnAsnLeuValAspLeuSerThrGlnIlePheCysHis 20
DB 185 CCCGCTGATGAAATGCGGGGCAAAACCTGCTGATCTTCGACGCAATCTTTGCCAT 244
QY 21 AsnAspTyrProGlnThrIleThrAspTyrValThrLeuGlnArgIleSerAlaTyrGly 40
DB 245 AACGATTAATCCGGAACATTAACAGACTATGTCACACTGCACAGAGCTCGCTATGGC 304
QY 41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
DB 305 GCGGTGTATATCTTATTTTCGCGGACCGT-AAATATAGTGCGAGTAGCTATCCATTTCT 363
QY 61 ThrThrSerGlnThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
DB 364 ACCACACGAGCAACGCGCGCTTGTATTAATTCAGAAAGCATTAAGCGGTGGCGGTG 423

QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
DB 424 GCGCTTATTTGACGCGCTGTGACGAGTGGCGGGGTGGCGATTAAGCTGCTCATTTA 483
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
DB 484 ATTGCGTGTCTTATTTTGGCAGACACCAACACTATATACGCGATGATTTCCAGTTGTG 543
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
DB 544 TGGATATTTTACGCGCAATATATGATGTGTGCTGCTGCTGCGGCTCGATGTTCTGCT 603
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 604 CGTATGTCACTTACTGTGCGGACTACCTGTTCAGTGCATTCCTTACCGTT 663
QY 161 Tyr 161
DB 664 TAT 666

Search completed: November 28, 2002, 19:59:12
Job time : 2666 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:

November 28, 2002, 19:09:21 : Search time 49 Seconds
(without alignments)
1007.652 Million cell updates/sec

Title:

US-09-900-575-29_COPY_26_186

Perfect score:

848

Sequence: 1 PNVNNGNLVVDISTQIFCH.....DVFVTLDPYRGSVPIPLTV 161

Scoring table:

BLOSUM62
Xgapop 10.0 , Ygapext 0.5
Ygapop 10.0 , Ygapext 0.5
Xgapop 6.0 , Ygapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

MODEL=frame_plus_p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US0900575/runat.22112002.130709.4580/app.query.fasta.1.327
-DB=Issued_Patents_NA -OEMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonum62 -TRANS=human40.cdl
-LIST=45 -DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=45
-MODE=LOCAL -OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0900575 -CGN1_1_17=etunal.22112002.130709.4580 -NCPU=6 -ICPU=3
-NO.XLPHY -NO.MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA.*
1: /cgn2_6/prodata/1/lna/5A.COMB.seq.*
2: /cgn2_6/prodata/1/lna/5B.COMB.seq.*
3: /cgn2_6/prodata/1/lna/6A.COMB.seq.*
4: /cgn2_6/prodata/1/lna/6B.COMB.seq.*
5: /cgn2_6/prodata/1/lna/6C.COMB.seq.*
6: /cgn2_6/prodata/1/lna/backfillseq1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96.5	11.4	30549	US-09-134-001C-322	Sequence 322, App
2	84.5	10.0	4403765	US-09-103-840A-2	Sequence 2, Appl
3	79	9.3	2717	US-07-723-002C-1	Sequence 1, Appl
4	78	9.2	537	US-09-134-001C-2157	Sequence 2157, Ap
5	77.5	9.1	1777	US-08-123-781-54	Sequence 54, Appl
6	77.5	9.1	1777	US-08-630-918-54	Sequence 54, Appl
7	77.5	9.1	1777	US-09-004-422-54	Sequence 54, Appl
8	77.5	9.1	1777	US-08-309-341-3	Sequence 3, Appl
9	77.5	9.1	2002	US-08-608-267-3	Sequence 3, Appl
10	77.5	9.1	2002	US-08-608-452-3	Sequence 3, Appl
11	77.5	9.1	2002	US-08-608-224-3	Sequence 3, Appl
12	77.5	9.1	2002	US-08-967-149-3	Sequence 3, Appl

13	75.5	8.9	1916	US-08-714-918-96	Sequence 96, Appl
14	75.5	8.9	1916	US-09-265-315-96	Sequence 96, Appl
15	75.5	8.9	1916	US-09-265-315-96	Sequence 96, Appl
16	75.5	8.9	1916	US-09-265-315-96	Sequence 96, Appl
17	75.5	8.8	420	US-09-661-638-65	Sequence 65, Appl
18	75	8.8	957	US-08-684-862-11	Sequence 11, Appl
19	75	8.8	5319	US-08-169-927-1	Sequence 1, Appl
20	74.5	8.8	7866	US-09-453-702B-102	Sequence 102, App
21	74	8.7	17056	US-09-245-041-3	Sequence 3, Appl
22	73	8.6	4162	US-09-453-702B-259	Sequence 259, App
23	72.5	8.5	1636	US-08-714-402-1	Sequence 1, Appl
24	72	8.5	3531	US-08-714-402-1	Sequence 1, Appl
25	72	8.5	3688	US-09-327-536-1	Sequence 39, Appl
26	72	8.5	25165	US-09-453-702B-39	Sequence 1, Appl
27	71.5	8.4	1843	US-07-918-023-1	Sequence 1, Appl
28	71.5	8.4	8855	US-08-542-003-1	Sequence 1, Appl
29	71.5	8.4	8855	US-08-322-760A-1	Sequence 1, Appl
30	71.5	8.4	8855	US-09-336-949-1	Sequence 1, Appl
31	71.5	8.4	9100	US-08-743-637B-27	Sequence 27, Appl
32	71.5	8.4	9100	US-08-526-840B-27	Sequence 27, Appl
33	71.5	8.4	9421	US-08-370-319C-2	Sequence 2, Appl
34	71.5	8.4	9421	US-09-224-834-2	Sequence 1, Appl
35	71	8.4	786	US-09-110-959A-1	Sequence 263, App
36	71	8.4	1463	US-09-071-035-263	Sequence 257, App
37	71	8.4	4977	US-09-071-035-261	Sequence 261, App
38	71	8.4	4977	US-09-071-035-261	Sequence 265, App
39	71	8.4	32768	US-08-961-527-71	Sequence 71, Appl
40	70.5	8.3	1870	US-09-295-186-9	Sequence 9, Appl
41	69.5	8.2	9319	US-08-976-259-85	Sequence 85, Appl
42	69.5	8.2	22671	US-08-976-259-85	Sequence 14, Appl
43	69.5	8.2	1408	US-08-712-072C-1	Sequence 1, Appl
44	69	8.1	1860	US-09-620-412C-308	Sequence 308, App
45	69	8.1	1860	US-09-620-412C-308	Sequence 308, App

ALIGNMENTS

RESULT 1
US-09-134-001C-322
Sequence 322, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 322
LENGTH: 30549
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-322

Alignment Scores:

Pred. No.: 1.21
Score: 96.50
Percent Similarity: 45.00%
Best Local Similarity: 27.22%
Query Match: 11.38%
Length: 30549
Matches: 49
Conservative: 32
Mismatches: 62
Indels: 37
Gaps: 11

US-09-900-575-29_COPY_26_186 (1-161) x US-09-134-001C-322 (1-30549)

Qy 2 ValValAsnValIGlyGlnAsnLeuValValAsnLeu
DB 5770 GTTCAATGTTCCACAAATGACAGTCTATATCATCAATGAATGTAATT 5829

```
OY 14 ---SerthrglnllephesCysHlsasnAspTyrProGlnThrIleThrAspTyrValThr 32
      :::::
      5830 CCTAATACAAATACAACTATAGATTCTAATGTTAGCTACGTACAAATCAAGCAGCT 5889
OY 33 LeuGlnArgGly-----SerAlaTyrGlyGlyValLeuSerAsnProSerGlyThrVal 50
      |||
      5890 CTACCAACCGGAATATATCTAGCTAAACCTCAATGACAAATATATATGACCTACACT--- 5946
OY 51 LysTyrSerGlySerSerTyrProPheProThrThrSerGlnThrProAlaValValTyr 70
      |||
      5947 AAACAAATAGTATGATGATGCTTCAATACAACTGACGATATAGT-----GTTT 6000
OY 71 AsnSerArgThrAspLysProTyrProValAlaLeuTyrIleThrProValSerSerAla 90
      :::::
      6001 TCAGAAACAGCTGATCAA-----GTAAGGTGTTACC 6030
OY 91 GlyGlyLeuValIleValAlaGlySerIleuIleAlaValLeuIleuArgGlnThrAsn 110
      |||
      6031 GCTGCACGCAAGCTAAATATGAT-----GCTATTAATATATTAAGGTACA--- 6078
OY 111 AsnTyrAsnSerAspAspPhe---GlnPheValTyrAsnIleTyrAlaAsnAsnAspVal 129
      |||
      6079 AACTATATTTTAAATGACTTCATAGTATATACAGCCCATTTCTACTCTT 6138
OY 130 Val-----ValProThr-----GlyGlyCysAspValSerAlaArgAspVal 143
      |||
      6139 ACATGGAACGAGAGCCTAATATAGTTGAAAAACAACATCGTACTACACAAACAACTGTT 6198
OY 144 ThrValThrIleuProAspTyrArgGly-----SerValProIleuProLeuThrValTyr 161
      |||
      6199 ACAGTTACTCTACTAATCATCATCAAGCTACGAGACTGATATATTCAAATACATCTAT 6258
```

RESULT 2

US-09-103-840A-2/c
Sequence 2, Application US/09103840A

Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:

Pred. No.:	5.75e+04	Length:	4403765
Score:	84.50	Matches:	42
Percent Similarity:	37.76%	Conservative:	12
Best Local Similarity:	29.37%	Mismatches:	72
Query Match:	9.96%	Indels:	17
DB:	4	Gaps:	5

US-09-900-575-29_COPY_26_186 (1-161) x US-09-103-840A-2 (1-4403765)

```
OY 27 IleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPro 46
      |||
      586389 ATCTCCAGCAACTGACGAGAAAGTCGGGCCCATTTGGACACAGCGCTGACACACC 586330
OY 47 SerGlyThrValTyrSerGlySerTyrProPheProThrThrSerGlnThrPro 66
```

```
DB 586329 GCGGATGGGTGAGCCCTCCGGGCTACGACGACACACTGCCGCGCCACCTCG 586270
      :::::
OY 67 ---ArgValValTyrAsnSerArgThrAspLysProTyrProValAlaLeuTyrIleThr 85
      |||
      586269 ATCAGCCCATCTGTGTGACGTGCGGCC---AGACCTGCTGACGAGAGTGGGAGACA 586213
OY 86 ProValSerSerAla-----GlyGlyLeuValIleValAlaGly----- 98
      |||
      586212 CCGAGCCGCTGCTCAGGCTCGCTGCGCGGCGAGGCTGACGCCCGCGGAAAGACCGC 586153
OY 99 -----SerIleuIleAlaValLeuIleuArgGlnThrAsnAsnTyrAsnSer 114
      |||
      586152 TCAGAGATGGCGTCGCGGATCGTCGCGATCCGCTCATCCACCGCTGCCGATCCGCGC 586093
OY 115 AspAspPheGlnPheValTyrAsnIleTyrAlaAsnAsnAspVal-----Val 130
      |||
      586092 GCCTGAAGAAGACCTTGACTGCTTCTGCTCAACCAATATAGTGGCCGCAATCGCACT 586033
OY 131 ValProThrGlyGlyCysAspValSerAlaArgAspValThrValThrIleuProAspTyr 150
      |||
      586032 CAGCCACGACGAGAGTGGCCGACGTGAGACGATGAACGCGGCGCGCTGCGCGGTGG 585973
OY 151 ArgGlySer 153
      |||
      585972 CGCGGTGCA 585964
```

RESULT 3

US-07-723-002C-1
Sequence 1, Application US/07723002C

Patent No. 5447862
GENERAL INFORMATION:

APPLICANT: Heim, Jutta
APPLICANT: Meyhack, Bernd
APPLICANT: Gysler, Christof
APPLICANT: Visser, Jacob
APPLICANT: Kester, Hermanus Cornelis Maria
TITLE OF INVENTION: No. 5447862el Expression System
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/723,002C
FILING DATE: 28-JUN-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8702475
FILING DATE: 04-FEB-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 150,880
FILING DATE: 29-JAN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8818046.8
FILING DATE: 28-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8914666.6
FILING DATE: 26-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 384,898
FILING DATE: 24-JUL-1989
ATTORNEY/AGENT INFORMATION:
NAME: Foley, Shawn P.
REGISTRATION NUMBER: 33,071

```

REFERENCE/DOCKET NUMBER: 4-16317/+CIP
TELEPHONE: (919)541-8615
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2717 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Aspergillus niger N756
IMMEDIATE SOURCE:
CLONE: Plasmid pCG3B11 (DSM 3916)
FEATURE:
NAME/KEY: CDS
LOCATION: join(689..889, 955..1098, 1161..1286, 1350..1445,
LOCATION: 1503..2034)
OTHER INFORMATION: /transl_except- (pos: 1096 .. 1098, aa: Tyr)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 689..746
FEATURE:
NAME/KEY: exon
LOCATION: 747..889
FEATURE:
NAME/KEY: Intron
LOCATION: 890..954
FEATURE:
NAME/KEY: exon
LOCATION: 955..1097
FEATURE:
NAME/KEY: Intron
LOCATION: 1098..1159
FEATURE:
NAME/KEY: exon
LOCATION: 1160..1286
FEATURE:
NAME/KEY: Intron
LOCATION: 1287..1349
FEATURE:
NAME/KEY: exon
LOCATION: 1350..1445
FEATURE:
NAME/KEY: Intron
LOCATION: 1446..1502
FEATURE:
NAME/KEY: exon
LOCATION: 1503..2054
FEATURE:
NAME/KEY: promoter
LOCATION: 1..688
FEATURE:
NAME/KEY: terminator
LOCATION: 2058..2717
US-07-723-002C-1

Alignment Scores:
Pred. No.: 5.42 Length: 2717
Score: 79.00 Matches: 38
Percent Similarity: 39.13% Conservative: 25
Best Local Similarity: 23.60% Mismatches: 72
Query Match: 9.32% Indels: 26
DB: 1 Gaps: 7

US-09-900-575-29_COPY_26_186 (1-161) x US-07-723-002C-1 (1-2717)
QY 9 LeuValValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr 28
DB 1533 CTCGGTACCGACGCCGACGCCGCTGCTCCATCCACCAACACTACATCAACGCGAGTCT 1592
QY 29 AspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyVal---LeuSerAsnPheSer 47

```

```

DB 1593 GACTACTCTGCTACTTCGAGCGCCACCACTACGAAACGTGACTTGAACGCTCAAC 1652
QY 48 GlyThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArg 67
DB 1653 GACAAAGTCACTTCAGAGGC---AACTACCTGTACAAGACCTCCGCCGCTGCCCAAG 1709
QY 68 Val-----ValTyrAsnSerArgThrAspLysProTyrProVal 80
DB 1710 GTCCAGACAAACACTTACCTCCATCTACCACTACTGAGGAACAACATCGGGCCAC 1769
QY 81 AlaLeuTyrIleThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
DB 1770 GCTTTCGAGATC-----GCTCCGGTGGCTACTCTCGCGGAGGTAATCAAC 1817
QY 101 IleAla---ValIleuIleuArgGlnThrAsnAsn-----TyrAsnSer 114
DB 1818 TTCTCCAACTGACGACACCGCTCTCGAGACCGACCTTCGAGGTGCTCTCTCTCTCT 1877
QY 115 AspAspPheGlnPheValTyrAsnIleTyrAlaAsnAsnAspValValValProThrGly 134
DB 1878 GACAGCGCTCTCTCCACCTGCGAGTCTTACATTGGCCGTTCTGCTTGCACCACTCAAC 1937
QY 135 GlyCysAspValSerAlaArgAspValThrVal-----ThrLeu 147
DB 1938 GCGGTGACCTCAACGCGACCTCCACACCGTCTCTCCAACTCAACGCGGACACCTTC 1997
QY 148 Pro 148
DB 1998 CCC 2000

RESULT 4
US-09-134-001C-2157
Sequence 2157, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 2157
LENGTH: 537
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2157

Alignment Scores:
Pred. No.: 0.663 Length: 537
Score: 78.00 Matches: 20
Percent Similarity: 48.44% Conservative: 11
Best Local Similarity: 31.25% Mismatches: 19
Query Match: 9.20% Indels: 14
DB: 4 Gaps: 2

US-09-900-575-29_COPY_26_186 (1-161) x US-09-134-001C-2157 (1-537)
QY 96 LysAlaGlySerLeuIleAlaValIleuArgGlnThrAsnAsnTyrAsnSerAsp 115
DB 100 AAAGGTGTCACGTGATTCATCTTGATTAATCAACATGAGATTCATGTGATTAATCA 159
QY 116 AspPheGlnPheVal-----TyrAsn 122
DB 160 CACATATGCAATTTTAAGATGATATGAAAGCTATTACACACCACTTAAGCTTTGGTTA 219
QY 123 IleTyrAlaAsnAsn---AspValValValProThrGlyGlyCysAspValSerAlaArg 141

```

Db 220 ACTGTCATCAATCACTAGATGTCATTATACACGGAGAGACGATATCTCACAAGA 279
QY 142 Aspyalthrval 145
|||||
Db 280 GATGTGACATA 291

RESULT 5
US-08-229-781-54
; Sequence 54, Application US/08229781
; Patent No. 5589174
; GENERAL INFORMATION:
; APPLICANT: Yoshinobu OKUNO et al.
; TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/229,781
; FILING DATE: April 19, 1994
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/054,016
; FILING DATE: April 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:

TELEX:
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1777 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to genomic RNA
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: A2/Aich1/2/68
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:

;; PUBLICATION INFORMATION:
;; AUTHORS:
;; TITLE:
;; JOURNAL:
;; VOLUME:
;; ISSUE:
;; PAGES:
;; DATE:
;; DOCUMENT NUMBER:
;; FILING DATE:
;; PUBLICATION DATE:
;; RELEVANT RESIDUES IN SEQ ID NO:
US-08-229-781-54

Alignment Scores:
Pred. No.: 4.47 Length: 1777
Score: 77.50 Matches: 44
Percent Similarity: 38.01% Conservative: 21
Best Local Similarity: 25.73% Mismatches: 67
Query Match: 9.14% Indels: 39
Gaps: 9
DB: 1

US-09-900-575-29_COPY_26_186 (1-161) x US-08-229-781-54 (1-1777)

QY 1 ProvalValasnValGlyGlnAsnLeuValValAspLeu-----SerThrGlnIle 17
||| :||| |||||
Db 304 CCTCATGTGATGTTTTCAAAATGACACATGGAGCTTTGCTGACGCAACAAAGCT 363
||| ||| |||||
QY 18 PheCysHlsAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnAsnArgLys 37
||| ||| |||||
Db 364 TTC--AGCAACTGTTACCTTATGATGTCAGATTAATTCCTTCCTTACGTCACACTAGTT 420
||| ||| |||||
QY 38 AlaTyrGlyGlyValLeuSer-----AsnPheSerGlyThrValLys 51
||| ||| |||
Db 421 CCTCGTCAGCAGCTCTGAGATTATCATCGAGGCTTTCACCTTGAGACTGGGCTCAG 480
||| ||| |||||
QY 52 TyrSerGlySerTyrProPheProThrThrSerGluThrProAlaGlyValTyrAsn 71
||| ||| |||||
Db 481 AATGGGGAGAACAAAT-----GCTTGCAAAAGGAGGAGCTGTTAGCGGTTTTTC 528
||| ||| |||||
QY 72 SerArgThrAspLysProTyrProValAlaLeuTyrLeuThrProValSerSerAlaGly 91
||| ||| |||||
Db 529 AGTAGACTGAAC-----TGG----- 543
||| ||| |||||
QY 92 GlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsn 111
||| ||| |||||
Db 544 -----TTGACCAATTCAGGAAGCACATATCCAGTGTGAGAGTACTATGCCAAACAAAT 597
||| ||| |||||
QY 112 TyrAsnSerAspPheGlnPheValTyrPasnIleTyr-----AlaAsnAsp 128
||| ||| |||||
Db 598 GACAAATTTTGACAAACTA---TACATTGGGGGATTCACACCCGAGCAGCAACAGAA 654
||| ||| |||||
QY 129 -----ValValValProThrGlyGlyCysAspValSerAlaArgAspValThr 144
||| ||| |||||
Db 655 CAACACAGCGCTATGTTCAACAGCATCAGAGAGTCACAGCTCTACACGAGGAAGCAG 714
||| ||| |||||
QY 145 ValThrLeuProAspTyrArgLysValPro 155
||| ||| |||||
Db 715 CAACATATATCCGAAATATCGGTCAGACC 747
||| ||| |||||

RESULT 6
US-08-630-918-54
; Sequence 54, Application US/08630918
; Patent No. 5631350
; GENERAL INFORMATION:
; APPLICANT: Yoshinobu OKUNO et al.
; TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.

COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,918
FILING DATE: April 5, 1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,781
FILING DATE: April 19, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/054,016
FILING DATE: April 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 1777 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to genomic RNA
ORIGINAL SOURCE:
ORGANISM: A2/Alchl/2/68
US-08-630-918-54

Alignment Scores:
Pred. No.: 4.47 Length: 1777
Score: 77.50 Matches: 44
Percent Similarity: 38.01% Conservative: 21
Best Local Similarity: 25.73% Mismatches: 67
Query Match: 9.14% Indels: 39
Gaps: 9

US-09-900-575-29_COPY_26_186 (1-161) x US-08-630-918-54 (1-1777)

QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeu-----SerThrGlnIle 17
Db 304 CCTCATGTGATGTTTTCAAATGACATGGGACCTTTCCTTGACACCAACAAAGCT 363
QY 18 PhcYsHtAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArGlySer 37
Db 364 TTC---AGCAACTGTTACCTTATGATGTCAGATTATGCTCCTTAGCTCCTAGT 420
QY 38 AlaTyrGlyValLeuSer-----AsnPhSerGlyThrValys 51
Db 421 GCCTCGTCAGCAGCTCGAGGTTTATCAGCAGAGGTTTCACTTGGAGCTGGGCTGACTCAG 480
QY 52 TyrSerGlySerTyrProPheProThrThrSerGluThrProArgValValTyrAsn 71
Db 481 AATGGGGAGCAAT-----GCTTGCAGAAAGGAGCTGTGCGGTTTTTTC 528
QY 72 SerArgThrAspLysProTrrProValAlaLeuTyrLeuThrProValSerSerAlaGly 91
Db 529 AGTAGACTGAC-----TGG----- 543
QY 92 GlyLeuValIleLeuAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsn 111
Db 544 -----TTGACCAATACAGAGACATATCCAGTGTCTAAGTACTATGCCAACAAT 597
QY 112 TyrAsnSerAspAspPheGlnPheValIleTyrPasnIleTyr-----AlaAsnAsnAsp 128
Db 598 GACAAATTTTGACAAACTA---TACATTGGGGGATTCCACACCGACGACGAAACAGAA 654

QY 129 -----ValValValProThrGlyGlycysAspValSerAlaArgAspValThr 144
Db 655 CAACAGCCCTGTATGTTTCAAGCATCAGGAGAGACTCACACTCTCTTACCAGAGAACCCAG 714
QY 145 ValThrLeuProAspTyrArgGlySerValPro 155
Db 715 CAACATATATCCGAAATATCGGCTCCAGACC 747

RESULT 7
US-09-004-422-54
Sequence 54, Application US/09004422
Patent No. 637070
GENERAL INFORMATION:
APPLICANT: Yoshinobu OKUNO et al.
TITLE OF INVENTION: POLYPEPTIDES FOR USE IN GENERATING
TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODIES (AS AMENDED)
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 203 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/004,422
FILING DATE: January 8, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/443,862
FILING DATE: May 22, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,781
FILING DATE: April 19, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/054,016
FILING DATE: April 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 1777 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to genomic RNA
HYPOTHETICAL:
ANTISENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: A2/Alchl/2/68
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:

```
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:

US-09-004-422-54

Alignment Scores:
Pred. No.: 4.47 Length: 1777
Score: 77.50 Matches: 44
Percent Similarity: 38.01% Conservative: 21
Best Local Similarity: 25.73% Mismatches: 67
Query Match: 9.14% Indels: 39
DB: Gaps: 9

US-09-900-575-29_COPY_26_186 (1-161) x US-09-004-422-54 (1-1777)
QY 1 ProValValAsnValGlyGlnAsnLeuValAlaSplu-----SerThrcGlnIle 17
DB 304 CCTCATGTTGATGTTTTCATAATAGACATGGACCTTTCGTTGACGCAAGCT 363
QY 18 PheCysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySer 37
DB 364 TTC---AGCAACTGTTACCTTGTGATGTGCAGATTATGCCCTTAGGTCACACTATT 420
QY 38 AlaTyrGlyGlyValLeuSer-----AsnPheserGlyThrValLys 51
DB 421 GCCTCGTCAGGACACTGTGAGTTTACACGACGAGGTTTCACCTGGACCTGGGCTCACTAG 480
QY 52 TyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsn 71
DB 481 AATGGGGGGAAGCAAT-----GCTTGCAAAAGGGGACCTGCTAGCGGTTTTC 528
QY 72 SerArgThrAspLysProTTPProValAlaLeuTyrLeuThrProValSerSerAlaGly 91
DB 529 AGTAGACTGAAC---TGG----- 543
QY 92 GlyLeuValIleLysAlaGlySerLeuIleAlaValIleLeuArgGlnThrAsnAsn 111
DB 544 -----TTGACCAATAGAGAGACATATCCAGTCTGCAAGCTATGCCAAACAAAT 597
QY 112 TyrAsnSerAspAspPheGlnPheValITrpaSnIleTyr-----AlaAsnAsnAsp 128
DB 598 GACAAATTTTGACAAACTA---TACATTGGGGGATTCACACCCGACGACGACCAAGAA 654
QY 129 -----ValValAlaProThrGlyGlyCysAspValSerAlaArgAspValThr 144
DB 655 CAAACCAACCTGTTATGTTTCAAGCATCAGGGAGACTCACAGTCTCTACAGAGAGAACCCAG 714
QY 145 ValThrLeuProAspTyrArgGlySerValPro 155
DB 715 CAAACTATATATCCGAATATATCGGGTCCAGACC 747

RESULT 8
US-08-309-341-3/C
; Sequence 3, Application US/08309341
```

```
Patent No. 5594119
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
APPLICANT: Thompson, Sheryl Ann
TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 55941190 No. 5594119disk of No. 5594119th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/309,341
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lowney, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247,000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2002 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Aspergillus niger
FEATURE:
NAME/KEY: Intron
LOCATION: 349..411
FEATURE:
NAME/KEY: CDS
LOCATION: join (348..412)

US-08-309-341-3

Alignment Scores:
Pred. No.: 5.33 Length: 2002
Score: 77.50 Matches: 41
Percent Similarity: 37.75% Conservative: 16
Best Local Similarity: 27.15% Mismatches: 52
Query Match: 9.14% Indels: 42
DB: Gaps: 6

US-09-900-575-29_COPY_26_186 (1-161) x US-08-309-341-3 (1-2002)
QY 45 AsnPheserGlyThrValLysTyrSerGlySerSer-----TyrProPheProThr 61
DB 1220 AACTATAGCAAGACTCAATCATAGACTGGACGAGGAGGAGCGCTTCCATGGCTGCG 1161
QY 62 ThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTTP-ProValAla 81
DB 1160 AGGAGCCCTCGTCCAAAGACAGCTGGGTAAACACCGTCACACAGCCATGGAGCGTAGT 1101
QY 81 aleuTyrLeuThr-----ProValSerSerAlaGlyGlyLeuValIleLys 96
DB 1100 ACTCGTACTAGAGAGAGACCGCTGCTGACACCGTTCGCAATAGAAAGGAGCTGACGTTGA 1041
QY 96 salaglySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAs 116
DB 1040 T-----GTTGGCGTTCTTGTGACACA-----AATCTCCGAAGCAAGA 1002
QY 116 pheGlnPheValITrpaSnIleTyrAlaAsnAsnAspValValValProThrGly----- 134
```


DB 1001 CGGGGATATAGTACAGCATAGATTACCGGCATGTGAAGTCTCTTGGCATACT 942
QY 135 -----GlyCysaspValSerAla----- 140
DB 941 CGGGGAATTGTTTGAAGAAGAGGTAAAGCAGCATAGACGTCTTGGCCAGCAGCAGG 882
QY 141 -----ArgaspValThrValThrLeuProAspTy 150
DB 881 TGTGCTGACAGCAGATTGCTGTAAGAGTAACCGACGCTTGACAGCGTGTCAAGAGA 822
QY 150 rArgGlySerValProIleProLeuThrVal 160
DB 821 TCACGAGCCGTTGGAGTTCACAGCGTAGTC 791

RESULT 9
US-08-608-267-3/C

; Sequence 3, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5688663disk of No. 5688663th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,267
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 349..411
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join (348..412)
; US-08-608-267-3

Alignment Scores:
Pred. No.: 5.33
Score: 77.50
Percent Similarity: 37.75%
Best Local Similarity: 27.15%
Query Match: 9.14%
Length: 2002
Matches: 41
Conservative: 16
Mismatch: 52
Indels: 42

DB: 1 Caps: 6

US-09-900-575-29_copy_26_186 (1-161) x US-08-608-267-3 (1-2002)

QY 45 AsnHeserGlyThrValLysTySerGlySerSer-----TyProPheProThr 61
DB 1220 AACTATAGCAAGACTCATATAGACGTCGACGAGAGAGGCGGTGTTCATGGCCGGC 1161
QY 62 ThrSerGlyThrProValValValTyAsnSerArgThrAspLysProTrp-ProValAl 81
DB 1160 AGAGAGCCCTGCTCCAAACACAGCTGGTAACACCGTCACACAGCCATGAGGAGGTAGT 1101
QY 81 AleuTyLeuThr-----ProValSerSerAlaGlyGlyLeuValIleTy 96
DB 1100 ACTGTACTAGTAGACACCGTCGCTGAGACCGTTGCCAATTAACAGGACTGAGGTGA 1041
QY 96 sAlaGlySerLeuIleValValLeuIleValGlnThrAsnAspTyAsnSerAsp 116
DB 1040 T-----GTTGCGCTTCTTGTGACACA-----AACTCCGAACAAAGA 1002
QY 116 pPheGlnPheValITrPAsnIleTyAlaAsnAspValValProThrGly----- 134
DB 1001 CGGGGATATAGTACAGCATAGATTACCGGCATGTGAAGTCTCTTGGCATACT 942
QY 135 -----GlyCysaspValSerAla----- 140
DB 941 CGGGGAATTGTTTGAAGAAGAGGTAAAGCAGCATAGACGTCTTGGCCAGCAGCAGG 882
QY 141 -----ArgaspValThrValThrLeuProAspTy 150
DB 881 TGTGCTGACAGCAGATTGCTGTAAGAGTAACCGACGCTTGACAGCGTGTCAAGAGA 822
QY 150 rArgGlySerValProIleProLeuThrVal 160
DB 821 TCACGAGCCGTTGGAGTTCACAGCGTAGTC 791

RESULT 10

US-08-608-452-3/C
; Sequence 3, Application US/08608452
; Patent No. 5693510
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5693510disk of No. 5693510th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,452
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

```
LENGTH: 2002 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Aspergillus niger
FEATURE:
NAME/KEY: intron
LOCATION: 349..411
FEATURE:
NAME/KEY: CDS
LOCATION: join (348..412)
US-08-608-452-3

Alignment Scores:
Pred. No.: 5.33 Length: 2002
Score: 77.50 Matches: 41
Percent Similarity: 37.75% Conservative: 16
Best Local Similarity: 27.15% Mismatches: 52
Query Match: 9.14% Indels: 42
Gaps: 6
DB: 1

US-09-900-575-29_COPY_26_186 (1-161) x US-08-608-452-3 (1-2002)
QY 45 AsnpheserGlyThrValIysTyrSerGlySerSer-----TyrProPheProThr 61
DB 1220 AACTATACAGACAGCATCATATACATGACGACGAGAGGCGCTTGTCCATGGCCCTGGC 1161
QY 62 ThrSerGluThrProArgValAlaValIyrsnserArgThrAspIysProTyr-ProValAl 81
DB 1160 AGAGGCGCTGCTGCACAGACAGCTGGGTAAACCGCTCACACAGCCATGGAGCGTGTGT 1101
QY 81 aleuTyrleuThr-----ProValSerSerAlaGlyGlyLeuValIleIly 96
DB 1100 ACTCGTACAGAGAGAGACCGCTGCTGACACCGCTTGCATTAAGAACGCTGCAGGTGA 1041
QY 96 salagIySerleuIleAlaValIleuArgInThrAsnTyrAsnSerAspAs 116
DB 1040 T-----GTTGCGCTTCTGTGTGAGACAA-----AATCTCCGAAGCAAGA 1002
QY 116 pheGlnPheValItrPasnIleTyrAlaAsnAspValAlaValIProThrGly----- 134
DB 1001 CGGGGATATAGTACACGACATGATTCACCGGCATGTGGAAGCTCTGCTGGCATACT 942
QY 135 -----GlyCysAspValSerAla----- 140
DB 941 CGGGGATTTGTTGAAGAAGGGGTAAAGCAAGCATAGCTCTTGCACGACGACACGG 882
QY 141 -----ArgAspValItrValIthrLeuProAspTy 150
DB 881 TGTGCGTGACAGACAGATTCTGTAAAGATTAACCGCATGTCAGCGCTGTCAAGAAAGA 822
QY 150 rArgIySerValIProIleProLeuThrVal 160
DB 821 TCACGAGACGCGTTGAGACTTCCAAAGCGTAGTC 791

RESULT 11
US-08-608-224-3/c
; Sequence 3, Application US/08608224
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 57053760 No. 5705376disk of No. 5705376th America, Inc.
; STREET: 405 Lexington Avenue, suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/608,224
FILING DATE: 28-FEB-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/309,341
FILING DATE: 20-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowmey, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247,000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2002 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Aspergillus niger
FEATURE:
NAME/KEY: intron
LOCATION: 349..411
FEATURE:
NAME/KEY: CDS
LOCATION: join (348..412)
US-08-608-224-3

Alignment Scores:
Pred. No.: 5.33 Length: 2002
Score: 77.50 Matches: 41
Percent Similarity: 37.75% Conservative: 16
Best Local Similarity: 27.15% Mismatches: 52
Query Match: 9.14% Indels: 42
Gaps: 6
DB: 1

US-09-900-575-29_COPY_26_186 (1-161) x US-08-608-224-3 (1-2002)
QY 45 AsnpheserGlyThrValIysTyrSerGlySerSer-----TyrProPheProThr 61
DB 1220 AACTATACAGACAGCATCATATACATGACGACGAGAGGCGCTTGTCCATGGCCCTGGC 1161
QY 62 ThrSerGluThrProArgValAlaValIyrsnserArgThrAspIysProTyr-ProValAl 81
DB 1160 AGAGGCGCTGCTGCACAGACAGCTGGGTAAACCGCTCACACAGCCATGGAGCGTGTGT 1101
QY 81 aleuTyrleuThr-----ProValSerSerAlaGlyGlyLeuValIleIly 96
DB 1100 ACTCGTACAGAGAGACCGCTGCTGACACCGCTTGCATTAAGAACGCTGCAGGTGA 1041
QY 96 salagIySerleuIleAlaValIleuArgInThrAsnTyrAsnSerAspAs 116
DB 1040 T-----GTTGCGCTTCTGTGTGAGACAA-----AATCTCCGAAGCAAGA 1002
QY 116 pheGlnPheValItrPasnIleTyrAlaAsnAspValAlaValIProThrGly----- 134
DB 1001 CGGGGATATAGTACACGACATGATTCACCGGCATGTGGAAGCTCTGCTGGCATACT 942
QY 135 -----GlyCysAspValSerAla----- 140
DB 941 CGGGGATTTGTTGAAGAAGGGGTAAAGCAAGCATAGCTCTTGCACGACGACACGG 882
QY 141 -----ArgAspValItrValIthrLeuProAspTy 150
DB 881 TGTGCGTGACAGACAGATTCTGTAAAGACTTAACGACGCTGTGTCAGAAAGA 822
```

QY 150 farglySerValProIleProLeuThrVal 160
DB 821 TCACGGACCGCTTGAGTTCCAAAGCCTAGTC 791

RESULT 12
US-08-967-149-3/C
Sequence 3, Application US/08967149
Patent No. 5939305
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
APPLICANT: Thompson, Sheryl Ann
TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
TITLE OF INVENTION: ASPERGILLUS NIGER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: No. 59393050 No. 5939305disk of No. 5939305th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,149
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,452
FILING DATE: 28-FEB-1996
APPLICATION NUMBER: US 08/309,341
FILING DATE: 20-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowrey, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247,000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2002 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Aspergillus niger
FEATURE:
NAME/KEY: Inttron
LOCATION: 349..411
FEATURE:
NAME/KEY: CDS
LOCATION: Join (348..412)
US-08-967-149-3

Alignment Scores:
Pred. No.: 5.33 Length: 2002
Score: 77.50 Matches: 41
Percent Similarity: 37.75% Conservative: 16
Best Local Similarity: 27.15% Mismatches: 52
Query Match: 9.14% Indels: 42
DB: 2 Gaps: 6

US-09-900-575-29_copy_26_186 (1-161) x US-08-967-149-3 (1-2002)

QY 45 AsnphSerGlyThrValysrGlySerSer-----TyrProPheProThr 61
DB 1220 MACTATAGCAAGACTCATCATAGACTGGCAGGAGGAGGCGCTTGCTCATGCGCTGCGC 1161

QY 62 ThrsrGluThrProArGValValTyrAsnSerArgThrAspLysProTyr-ProValAl 81
DB 1160 AGAGCCCTGCTCCAAAGACAGCTGGTACCCAGCCGTCACAGCAGCGCCAGCGTAGT 1101

QY 81 aleuTyrLeuThr-----ProValSerSerAlaGlyGlyLeuValIlely 96
DB 1100 ACTGCTAGCTAGTGAGACCGCTGCTGAGACCGTTGCCAATAGAACGAGCTCCAGGTGA 1041

QY 96 salaglySerLeuIleValIleuIleuIleuArGInThrAsnAsnTyrAsnSerAspAs 116
DB 1040 T-----GTTCCGCTTCTTGTGAGACAA-----AATCTCCGAGCAAGA 1002

QY 116 pPheGlnPheValTyrPasnIleTyrAlaAsnAsnAspValValProthGly----- 134
DB 1001 CGGGGATATAGTACGACGATAGATTACCGGCAATGTGAAGTCTGCTGCTGCTACT 942

QY 135 -----GlyCysAspValSerAla----- 140
DB 941 CGGGGATTTGTTGAGAGAGGTAAGCAAGCATAGACGCTCTGCCAGCAGCAAGC 882

QY 141 -----ArgAspValThrValThrLeuProAspTyr 150
DB 881 TGTGCTGACAGCAGAGTCTCTTAAGACTAACCGAGCTTGACAGCGCTGTCAAGGAGA 822

QY 150 farglySerValProIleProLeuThrVal 160
DB 821 TCACGGACCGCTTGAGTTCCAAAGCCTAGTC 791

RESULT 13
US-08-714-918-96
Sequence 96, Application US/08714918
Patent No. 6037123
GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ving
APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
TITLE OF INVENTION: TARGET GENES
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,918
FILING DATE: September 13, 1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Waibury, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440

```

TELEX 67-3510
INFORMATION FOR SEQ. ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 1916 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-714-918-96

Alignment Scores:
Pred. No.: 8.91 Length: 1916
Score: 75.50 Matches: 28
Percent Similarity: 41.23% Conservative: 19
Best Local Similarity: 24.56% Mismatches: 14
Query Match: 8.90% Indels: 23
DB: Gaps: 4

US-09-900-575-29_COPY_26_186 (1-161) x US-08-714-918-96 (1-1916)
QY 11 ValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr----- 28
DB 498 ATGGAATGAATACACGATTCAGTACAGTGAACATCTCTAATCGAAATGTTAAACAGCAATT 557
QY 29 AspTyrValIleThrLeuGlnArgGlySerAlaTyrGlyValLeuSerAsnPheSerGly 48
DB 558 GATTGATTAATTAATTAACAAATTCACAAAGTTGCTTATGGGTGACGTGTACCGCTATAACAAGAA 617
QY 49 ThrValIleTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgVal 68
DB 618 GATATTAAATTACAGACGACGACGCAATTCGATTAATTAATTAATGCTGAATAATCCT----- 671
QY 69 ValTyrAsnSerArgThrAspLysProTrrProValAlaLeuTyrLeuThrProValSer 88
DB 672 ---TACAAACAACTTATGCCATCACCGACAGTAAATGAGCAATATCTTGACACA----- 722
QY 89 SerAlaGlyGlyLeuValIleLysAla----- 97
DB 723 -----GGTGGATATGGTGCTTCGATACAGTACAGCATGTATACATTAATTATACGATACCG 776
QY 98 -----GlySerLeuIleAlaValLeuIleLeuArgGln 108
DB 777 CCATATTATGATTCGATCGATGGTACGAAATTAATCATCATCATGAA 818

RESULT 14
US-09-265-315-96
; Sequence 96, Application US/09265315
; Patent No. 6187541
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving J.
; APPLICANT: Malouin, Francois
; APPLICANT: Matlin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
; TITLE OF INVENTION: TARGET GENES
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:

```

[illegible]

;; TITLE OF INVENTION: TARGET GENES
;; NUMBER OF SEQUENCES: 111
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Lyon & Lyon
;; STREET: 633 West Fifth Street
;; CITY: Los Angeles
;; STATE: California
;; ZIP: 90071-2066
;; COUNTRY: U.S.A.
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
;; MEDIUM TYPE: Storage
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: IBM P.C. DOS 5.0
;; SOFTWARE: Word Perfect 5.1
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/265,315
;; FILING DATE: March 9, 1999
;; CLASSIFICATION: 435
;;
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: 08/714,918
;; FILING DATE: September 13, 1996
;; APPLICATION NUMBER: 60/009,102
;; FILING DATE: December 22, 1995
;; APPLICATION NUMBER: 60/003,798
;; FILING DATE: September 15, 1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Warburg, Richard J.
;; REGISTRATION NUMBER: 32,327
;; REFERENCE/DOCKET NUMBER: 240/247
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (213) 489-1600
;; TELEFAX: (213) 955-0440
;; TELEX: 67-3510
;; INFORMATION FOR SEQ ID NO: 96:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1916 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;;
US-09-265-315-96
;
Alignment Scores:
Pred. No.: 8 91 Length: 1916
Score: 75.50 Matches: 28
Percent Similarity: 41.23% Conservative: 19
Best Local Similarity: 24.56% Mismatches: 44
Query Match: 8.90% Indels: 23
DB: 4 Gaps: 4
;
US-09-900-575-29_COPY_26_186 (1-161) x US-09-265-315-96 (1-1916)
OY 11 ValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr----- 28
DB 498 ATGGAATGATACACGATATTCAGTAAACATCCCTAGCTAGTAAGGTAACAGCAATT 557
OY 29 AspTyrValThrLeuGlnArgGlySerAlaTyrGlyValLeuSerAsnPheSerGly 48
DB 558 GATTACTTAATTAATTACATTACAGTTGCTATGCTGACGTGACGTATTAACAAGAA 617
OY 49 ThrValValTyrSerGlySerSerTyrProPheProThrThrSerGluThrProAlaGVal 68
DB 618 GATATTAATTAATACAGACACCGCAATTGATTAAGATTATATCTGAATAATCTCT----- 671
OY 69 ValTyrAsnSerArgThrAspLysProTyrProValAlaLeuTyrLeuThrProValSer 88
DB 672 ---TACAGAAGACTTATGCTCATCACCAGGTAAATGAGCAATATCTTGACACA----- 722
OY 89 SerAlaGlyLeuValIleLysAla----- 97
DB 723 -----GCTGATATGTGTTGTCAGTAAGACTGCATGTTATTAATTAATACGATACCG 776

OY 98 -----GlySerLeuIleAlaValLeuIleLeuArgGln 108
DB 777 CCATATTAATGATTCATGCTGATGACGAAATTAATATCATATCAATGAA 818
RESULT 16
US-09-266-417-96
; Sequence 96, Application US/09266417
; Patent No. 6228588
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ying J.
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick R.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/266,417
; FILING DATE: March 9, 1999
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/714,918
; FILING DATE: September 13, 1996
; APPLICATION NUMBER: 60/009,102
; FILING DATE: December 22, 1995
; APPLICATION NUMBER: 60/003,798
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 240/248
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1916 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-266-417-96
;
Alignment Scores:
Pred. No.: 8 91 Length: 1916
Score: 75.50 Matches: 28
Percent Similarity: 41.23% Conservative: 19
Best Local Similarity: 24.56% Mismatches: 44
Query Match: 8.90% Indels: 23
DB: 4 Gaps: 4
;
US-09-900-575-29_COPY_26_186 (1-161) x US-09-266-417-96 (1-1916)
OY 11 ValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr----- 28
DB 498 ATGGAATGATACACGATATTCAGTAAACATCCCTAGCTAGTAAGGTAACAGCAATT 557

```

OY 29 AsptyrValThrLeuGlnArgGlySerAlaTyrGlyValLeuSerAsnPhseSergly 48
Db 558 GATTAGTAAATTAACATTAACAAGTTGCTATGGGTACGCTGTACCGTATTAACAGAA 617
OY 49 ThrValysrTyrSerglySerSeryrProPheProThrThrSergluThrProArgVal 68
Db 618 GATATTAAATTAACAGCAGCAGCAATGAAATTAGAAATTAATGCTGAAAAATCCT----- 671
OY 69 ValIyrAsnSerArgThrAspLysProTrrProValAlaLeuTyrLeuThrProValSer 88
Db 672 ---TACAGAACCTTATGCGCATCCAGTCAACGTAATAATGAGCAATATCTTGACCA----- 722
OY 89 SerAlaGlyValLeuValIleLeuSala----- 97
Db 723 -----GGGGGATATGCTGCTTCGATTAAGTACAGCATGTATACCAATATTATACGATACCG 776
OY 98 -----GlySerLeuIleAlaValLeuIleLeuArgin 108
Db 777 CCATATTATGATTCGATGATGAGCAAAATTAATCATCATCATGAA 818

RESULT 17
US-09-641-638-65
; Sequence 65, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bouguenere, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIOMIMETIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET 051CP1
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 65
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 364
; OTHER INFORMATION: 12-230-364 : polymorphic base C or T
; NAME/KEY: misc_binding
; LOCATION: 344..363
; OTHER INFORMATION: 12-230-364.misl, potential
; NAME/KEY: misc_binding
; LOCATION: 365..384
; OTHER INFORMATION: 12-230-364.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 1..20
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 401..420
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 352..376
; OTHER INFORMATION: 12-230-364 potential probe
US-09-641-638-65

Alignment Scores:
Pred. No.: 1.1 Length: 420
Score: 75.00 Matches: 31

```

```

Percent Similarity: 40.16% Conservative: 20
Best Local Similarity: 24.41% Mismatches: 58
Query Match: 8.84% Indels: 18
DB: 4 Gaps: 3

US-09-900-575-29_copy_26_186 (1-161) x US-09-641-638-65 (1-420)

OY 32 ThrLeuGlnArgGlySerAlaTyrGlyValLeuSerAsnPhseSerglyThrVallys 51
Db 26 TCTATTAATAAGGGGCAATGCTTCAAGAGATTTGTGTAGACTTCAGTCAAGTAAGCT 85
OY 52 TyrSerglySerSeryrProPheProThrThrSergluThrProArgValIleIyrAsn 71
Db 86 TGGGAGAGACCCAGC-----ACAACTCCAGGACAGGCGAGCGCTCTACAAA 133
OY 72 SerArgThrAspLysProTrrProValAlaLeuTyrLeuThrProValSergly 91
Db 134 TGTAGCTTCCCAATCCATTCATTCATCCATTTCTAGCCAGCGGTAGCACCTTTGA 193
OY 92 GlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsn 111
Db 194 TCT-----CTACGCCAGTCAATCT 214
OY 112 TyrAsnSerAspAspPheGlnPheValTrrAsnIleTyrAlaAsnAspValVal 131
Db 215 ACTCCAGTCACTATGCCCCCTCTGCTGATTTGTGAGGGGCAAGCAGCAAGGCG 274
OY 132 ProThrGlyGly---CysAspValSerAlaArgAspValThrValThrLeuProAspTyr 150
Db 275 CCAGAGGTGATATTGTCATCTCCCTGCTGCTATACCAACCTCAGTCACTGAGTTT 334
OY 151 ArgGlySerValProIlePro 157
Db 335 TTGGGTCCTTGTGCTGCT 355

RESULT 18
US-08-684-862-11
; Sequence 11, Application US/08684862
; Patent No. 5759541
; GENERAL INFORMATION:
; APPLICANT: Bach, Alfred
; APPLICANT: Hilden, Heinz
; APPLICANT: Blalogan, Siegfried
; TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 KB storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 5.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/684,862
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,705
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,040
; FILING DATE: 30-DEC-1992
; APPLICATION NUMBER: PCT/EP91/01361
; FILING DATE: 19-JUL-1991
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 957 base pairs

```

```

?      TYPE: nucleic acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: cDNA to mRNA
?      ORIGINAL SOURCE:
?      ORGANISM: Agkistrodon rhodostoma
?      FEATURE:
?      LOCATION: 210 to 911
?      OTHER INFORMATION: the coding region shown in (2)(1x)(B)
?      OTHER INFORMATION: codes for the protein of SEQ ID NO: 6
US-08-664-062-11

```

```
||||| :|||
Db 3220 -----GGTTAGGTTTGAAGATGGTAGT-----CCAAAGTTAAACAA 3258
Qy 109 -----ThrasnAsnTyraSerAspAspPheGlnPheValITrpAsnIleTyraLe 125
Db 3259 GTGACATTACTACAGATTATACAACTTAGGTAGTATTATGGCAAAATTAATGACAAATT 3318
Qy 126 AsnAsnAspValValProThrGlyGlyCysAspValSerAlaArgAspValThrVal 145
Db 3319 AATGATGATGACTCTTACTACAGAGGTATACAGAGGACAGATTGTCGCTAAATTT 3378
Qy 146 ThrLeuProAspTyraGlySerValProIle 156
Db 3379 ACTCTTGGAGTGTAAAGCTAACGCTAACGCTA 3411

RESULT 20
US-09-453-702B-102
; Sequence 102, Application US/09453702B
; Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESSES:
ADDRESS: Charles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296,95017
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 7886
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-09-453-702B-102

Alignment Scores:
Pred. No.: 95.8 Length: 7886
Score: 74.50 Matches: 27
Percent Similarity: 42.86% Conservative: 30
Best Local Similarity: 20.30% Mismatches: 57
Query Match: 8.79% Indels: 19
DB: Gaps: 3

US-09-900-575-29_COPY_26_186 (1-161) x US-09-453-702B-102 (1-7886)
```

```
Qy 3 ValAsnValGlyGlnAsnLeuValAlaAspLeuSerThrGlnIlePheCysHisAsnAsp 22
Db 7308 GTCCATTGTGGCAATACGAAGATCGATGATACGGACCACTCTTCAATTAATGGCGTC 7367
Qy 23 TyrProGluThrIleThrAspTyraValThrLeuGlnArgGlySerAlaTyrglyVal 42
Db 7368 TTGTGTGGCGGAGTACGATGATATATGACCCACGCTAACGCTAGC-----7412
Qy 43 LeuSerAsnPheSerGlyThrValIlystYrSerGlySerSerTyProPheProThr 62
Db 7413 -----TTCCGTGCGCAATTTGCGACATACGCGGCGCATGAACCCGACGACGAAA 7463
Qy 63 SerGluThrProArgValValTyraSerArgThrAspLysProTrpProValAlaLeu 82
Db 7464 ACACCGACGGCGCGCT-----AATGATTTCAATCCTGACCGCTG 7502
Qy 83 TyrLeuThrProValSerSerAlaGlyGlyLeuValIleLeuAlaGlySerIleLeu 102
Db 7503 TTGTGTGGCGGATTTCCAGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7562
Qy 103 ValLeuIleLeuArgGlnThrAsn-----AsnTyraSerAspAspPheGln 118
Db 7563 ATGATCGTCTCATTTTGTGCTGATATACGTTCTCCAAAGTAAAGTAAAGTATTATTAT 7622
Qy 119 PheValITrpAsnIleTyraAsnAsnAspValValVal 131
Db 7623 CTTATTTCATGGTGCATATTTCTGATATGAAATGAGAGATGCTG 7661

RESULT 21
US-09-245-041-3/c
; Sequence 3, Application US/09245041
; Patent No. 6274339
GENERAL INFORMATION:
APPLICANT: Moore, K.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
OF BODY WEIGHT DISORDERS INCLUDING OBESITY
FILE REFERENCE: 7853-136
CURRENT APPLICATION NUMBER: US/09/245,041
CURRENT FILING DATE: 1999-02-05
EARLIER APPLICATION NUMBER: 60/093,630
EARLIER FILING DATE: 1998-07-21
EARLIER APPLICATION NUMBER: 60/104,978
EARLIER FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 131
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 17056
TYPE: DNA
ORGANISM: Mus musculus
US-09-245-041-3

Alignment Scores:
Pred. No.: 345 Length: 17056
Score: 74.00 Matches: 30
Percent Similarity: 43.56% Conservative: 14
Best Local Similarity: 29.70% Mismatches: 38
Query Match: 8.73% Indels: 19
DB: Gaps: 3

US-09-900-575-29_COPY_26_186 (1-161) x US-09-245-041-3 (1-17056)

Qy 2 ValValAsnValGlyGlnAsnLeuValAlaAspLeuSerThrGlnIlePheCysHisAsn 21
Db 9975 GTGCTCAATCTCCACCAAGCTGATCTTGACCTTTGCAATCTTCTTC-----9925
Qy 22 AspTyProGluThrIleThrAspTyraValThrLeuGlnArgGlySerAlaTyrglyVal 41
Db 9924 -----CTTCTACATATTTTGAACCTTGAGAGGAGCTATAGCTTGGAGG 9880
Qy 41 yValLeuSerAsnPheSerGlyThrValIlystYrSerGlySerSerTyProPheProTh 61
Db 9879 AGAGAGATACACGATTTGGGAGTACGTAGTATGATGCTTAATCTTATGATGCTTGATT 9820
```



```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-08-908-643C-1

```

```

Alignment Scores:
Pred. No.: 16.8 Length: 1636
Score: 72.50 Matches: 39
Percent Similarity: 38.51% Conservative: 23
Best Local Similarity: 24.22% Mismatches: 60
Query Match: 8.55% Indels: 40
DB: Gaps: 6

```

```

US-09-900-575-29_COPY_26_186 (1-161) x US-08-908-643C-1 (1-1636)

```

```

QY 19 CysHisAsnAspTyrProGluThrIleThrAspTyrValIThrLeuGlnArgGlySerAla 38
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 340 TGTCATGT-GACTATTAAGAAAGCTTAACAGGCTGATGCTCCAGCTAGAAAGTTGATA 398
QY 39 TyrGlyGlyValLeuSerAsnPheserGlyThrValIlystYrSerGlySerSerTyrPro 58
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 399 TAC-----TTCCTGCTTAACCTT-----TGAAACCAACGAGCTGCC 437
QY 59 PheProThrSerGluThrProArgValValIlyrAsnSerArgThrAspLysProTrrp 78
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 438 TTCAAACTATTCTCGAGCACTCGTATGTTTACAAAGATGTCAGAAACTGAG--- 494
QY 79 ProValAlaLeuTyrLeuThrProValIserSerAlaGly-GlyLeuValIleYsAlaGl 98
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 495 -----GGACTGTTCTGTGACTTAATGCTTGAGGCTAGCCTTTCCTATTTCCTCCA 548
QY 98 YserLeuIleAlaValIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGl 118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 549 CGAAGCTGGCTTTAAGTGCTGTAAAGACATTAAGAGATTTCAG-----GA 596
QY 118 nPheValITpAsnIleTyrAlaAsnAspValValIValProThnGlyGlyCysAsp-- 137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 597 TATCTTAATGAGCCACACAGAAAGCAATGATTAATGTTGTGTGTCAGAGTT 656
QY 138 -----ValSerAlaArgAspVal----- 143
Db 657 CCTCTACAAGCTGAAGGGTACGACGACGAGCTGCAAGACATTCATTAATCTAGTGA 716
QY 144 -----ThrValIThrLeuProAspTyrArgGlySerVal 154
Db 717 TCTTTCAATGACCACTACTTGAGAGACAATGTCACGCCCTGACTATATGAAAAATGT 776
QY 154 1 154
Db 777 C 777

```

```

RESULT 24
US-08-714-402-1
; Sequence 1, Application US/08714402
; Patent No. 5910441
; GENERAL INFORMATION:
; APPLICANT: ROCHA, Claudia
; APPLICANT: FISCHETTI, Vincent A.
; TITLE OF INVENTION: FIBRONECTIN AND FIBRINOGEN BINDING
; TITLE OF INVENTION: PROTEIN FROM GROUP A STREPTOCOCCI
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,402
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm R.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 016921-097
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3531 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-714-402-1

```

```

Alignment Scores:
Pred. No.: 60.4 Length: 3531
Score: 72.00 Matches: 29
Percent Similarity: 37.21% Conservative: 19
Best Local Similarity: 22.48% Mismatches: 31
Query Match: 8.49% Indels: 50
DB: Gaps: 7

```

```

US-09-900-575-29_COPY_26_186 (1-161) x US-08-714-402-1 (1-3531)

```

```

QY 26 ThrIleThrAspTyrVal----- 31
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 940 ACTTTCACAGATTATATGCGGGTTAGATAAAGTCCAGTTGCTGAGAAATTGAGCTTA 999
QY 32 -----ThrLeuGlnArgGlySerAlaTyrGlyValIleuSerAsn 45
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1000 TTCCTAGAGAAATGAAGAGTGTGAAATGCTAGT-----ATCTCAAT 1044
QY 46 PheSerGlyThrValIlystYrSerGlySerSerTyrProPheProThrThrSerGluThr 65
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1045 TTTAAGATACCAATGAGTGGCGAGAGATCACCTAT-----AAGGAACG 1089
QY 66 ProArgValValIlyrAsnSerArgThrAspLysProTrrpProValAlaLeuTyrLeuThr 85
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1090 GTTAATGTTCTTAAGAAATGAGACGACTAAAGAA-----AGCATTAATTAATCT 1140
QY 86 --ProValSerSerAlaGlyGlyLeuValIleYsAlaGlySerLeuIleAlaValLeu 104
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 AATGATTGACCAATGTCGGGAGGTAT----- 1170
QY 105 IleuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheValITpAsnIleTyr 124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1171 -----CAAGCTACCAACACGAAAGCGAGAAATTTGCTGTATGTTAT 1215
QY 125 AlaAsn---AsnAspValValAlaPro 132
Db 1216 GTCAATCCAAACGTAACCAATATTCCT 1242

```

```

RESULT 25
US-09-327-536-1
; Sequence 1, Application US/09327536
; Patent No. 635477
; GENERAL INFORMATION:
; APPLICANT: ROCHA, Claudia
; APPLICANT: FISCHETTI, Vincent A.
; TITLE OF INVENTION: FIBRONECTIN AND FIBRINOGEN BINDING
; TITLE OF INVENTION: STREPTOCOCCI
; FILE REFERENCE: 022927-008
; CURRENT APPLICATION NUMBER: US/09/327,536
; CURRENT FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: US 08/714,402
; PRIOR FILING DATE: 1996-09-16

```

NUMBER OF SEQ ID NOS: 2
 SOFTWARE: Patentln Ver. 2.0
 SEQ ID NO 1
 LENGTH: 3698
 TYPE: DNA
 ORGANISM: SFFBP gene
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (196)..(3681)
 US-09-327-536-1

Alignment Scores:
 Pred. No.: 64.7 Length: 3698
 Score: 72.00 Matches: 29
 Percent Similarity: 37.21% Conservative: 19
 Best Local Similarity: 22.48% Mismatches: 31
 Query Match: 8.49% Indels: 50
 DB: 4 Gaps: 7

US-09-900-575-29_COPY_26_186 (1-161) x US-09-327-536-1 (1-3698)

OY 26 Thrilethraspyrval----- 31
 Db 940 ACTTTCACAGATTATATCGCGGTTAGATAAAGTCCAGTGTCTGCGAATTGAGCTTA 999
 OY 32 -----Thleuglnargllyseralatyrglyvalleuserasn 45
 Db 1000 TTCTAGAGAAATAGAGAGAGGAGTGTGGAAATACCTAGT-----ATCTCAAT 1044
 OY 46 Pheserglythrvallytyrserglyserserlyrprophethrthrserglutrh 65
 Db 1045 TTTAAGATACCATAGTGGGAGAGATCCTAT-----AAGGACG 1089
 OY 66 Proargvallytyrasnserargthrasyrprotrprovalalaletyrlleuthr 85
 Db 1090 GTTATGTTCTTATGAAATAGAGACGACCTAAGAA-----ACCATTAATTAAT 1140
 OY 86 ---Provalseralaglyglyleuvallelysalaglyserleullealavalleu 104
 Db 1141 AATGATGACCAATGTCGGGAGATAT----- 1170
 OY 105 lleuarglthrasnserargthrasyrprotrprovalalaletyrlleuthr 124
 Db 1171 -----GAAAGCTACACACGGAATTTGCTGCTGTTAT 1215
 OY 125 Alasn---asnasyrvalvalpro 132
 Db 1216 GTCATCCAAACGCTACCAATATTCCT 1242

RESULT 26
 US-09-453-702B-39
 Sequence 39, Application US/09453702B
 Patent No. 6365723
 GENERAL INFORMATION:
 APPLICANT: Blattner, Frederick R.
 Burland, Valerie
 Porne, Nicole T.
 Plunkett, Guy
 Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
 NUMBER OF SEQUENCES: 265
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Quarles & Brady
 STREET: 1 South Plinckney Street
 CITY: Madison
 STATE: WI
 COUNTRY: US
 ZIP: 53701-2113
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/453,702B
 FILING DATE: 03-Dec-1999
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/110,955
 FILING DATE: 04-Dec-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J.
 REGISTRATION NUMBER: 27386
 REFERENCE/DOCKET NUMBER: 960296,95017
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (608) 251-5000
 TELEFAX: (608) 251-9166
 INFORMATION FOR SEQ ID NO: 39:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 25165
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 39:
 US-09-453-702B-39

Alignment Scores:
 Pred. No.: 1.09e+03 Length: 25165
 Score: 72.00 Matches: 39
 Percent Similarity: 37.84% Conservative: 17
 Best Local Similarity: 26.35% Mismatches: 63
 Query Match: 8.49% Indels: 29
 DB: 4 Gaps: 5

US-09-900-575-29_COPY_26_186 (1-161) x US-09-453-702B-39 (1-25165)

OY 22 Aspyrprogluthrlethraspyrvalthrleuglnargllyseralatyrgly 41
 Db 20797 GATGTGGAGCGCGCGGTGATGCTGCTGCTCAACAGTCCGCGACCTGTC----- 20850
 OY 42 Valleserasnpheserlythrvallytyrserglyserserlyrprophethr 61
 Db 20851 ---CTGAGCGGACACCAACGAGGAGCTAC---GGCATTCCTACCTACCTCGGCATC 20904
 OY 62 Thrserglutthrproargvallytyrasnserargthrasyrprotrprovalala 81
 Db 20905 ACCGATCCCAAGCGCATTAACCTTAATACGCAATACCGTAGAG----- 20949
 OY 82 leutyrlleuthrprovalseralaglyglyleuvallelysalaglyserleulle 101
 Db 20950 -----CTGTACAACTCGCGGAGAGTCTGCTC 20976
 OY 102 Alavalleulleuarglthrasnserargthrasyrprotrprovalala 121
 Db 20977 GCCAGCAATATCATACCCGCGGTATGCGGTAAACACCTCGCGCTGCTC 21036
 OY 122 Asniletyr-----Alaasnasnasyrvalvalprothrthrglygly 135
 Db 21037 AACTCTACGCGCTGAATCCGCGAGAACCAACGCGGTGCTGATCAATCCACGCG 21096
 OY 136 Cysaspyrvalseralargasyrvalthrval-Thleuthrproasp-----Tyaraglyse 153
 Db 21097 ACCACCGCGCAATATGACGCGGTCAACACGAGCTGGCGGTTACAGCGCAC 21156
 OY 153 rvalproleuthrval 160
 Db 21157 GATGCCACTCACGCTTACGATC 21178

RESULT 27
 US-07-918-023-1
 Sequence 1, Application US/07918023
 Patent No. 5427934
 GENERAL INFORMATION:
 APPLICANT: Zimmermann, Thomas
 Robins, Karen

```

APPLICANT: Birch, Owen
APPLICANT: Bohlen, Elisabeth
TITLE OF INVENTION: GENETIC ENGINEERING PROCESS FOR THE
TITLE OF INVENTION: PRODUCTION OF 5-(+)-2,2-DIMETHYLCYCLOPROPANECARBOXAMIDE
TITLE OF INVENTION: BY MICROORGANISMS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fisher, Christen & Sabol
STREET: 2000 M Street, N.W., Suite 590
City: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/918,023
FILING DATE: 19920724
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2247/91
FILING DATE: 26-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Marsh, Virgil H.
REGISTRATION NUMBER: 23,083
REFERENCE/DOCKET NUMBER: LP 1521
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2000
TELEFAX: (202) 659-2015
TELEX: 248748
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1843 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Comamonas acidovorans
STRAIN: A:18
FEATURE:
NAME/KEY: misc_feature
LOCATION: 289..1566
PUBLICATION INFORMATION:
DOCUMENT NUMBER: IE 9224406
FILING DATE: 23-JUL-1992
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1843
PUBLICATION INFORMATION:
DOCUMENT NUMBER: JP 198717/92
FILING DATE: 24-JUL-1992
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1843
PUBLICATION INFORMATION:
DOCUMENT NUMBER: PL P 295408
FILING DATE: 24-JUL-1992
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1843
PUBLICATION INFORMATION:
DOCUMENT NUMBER: RO 92-01033
FILING DATE: 24-JUL-1992
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1843
PUBLICATION INFORMATION:
DOCUMENT NUMBER: SU UNKNOWN
FILING DATE: 24-JUL-1992
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1843
PUBLICATION INFORMATION:
DOCUMENT NUMBER: CS PV2323-92
FILING DATE: 24-JUL-1992
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1843
PUBLICATION INFORMATION:
DOCUMENT NUMBER: HU P9202439
FILING DATE: 24-JUL-1992

```

```

1      RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1843
2      US-07-918-023-1
3
4      Alignment Scores:
5      Pred. No.:      26.8      Length:      1843
6      Score:          71.50      Matches:      22
7      Percent Similarity: 42.67%      Conservative: 10
8      Best Local Similarity: 29.33%      Mismatches: 40
9      Query Match:      8.43%      Indels:      3
10     DB:                1      Gaps:          1
11
12     US-09-900-575-29_COPY_26_186 (1-161) x US-07-918-023-1 (1-1843)
13
14     QY      24      ProgluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyValLeu 43
15                ||| |||      ||||||||| ||| :|||
16     Db      1421     CCCCACACGGCCTTGGGGCCCCACGCTGCAACGCATGCGCATTTGGCTCCGATCC 1480
17
18     QY      44      SerAsnProSerGlyThrValLysTyrSerGlySerSerTyrProPheThrIleSer 63
19                ||||||| :||| ||| :||| ||| |||
20     Db      1481     GAGCGTCTCGGGCGCATGCGCTAC-----ACCTGCCGTTGCAGCTCACGGCC 1531
21
22     QY      64      GluThrProArgValValTyrAsnSerArgTyrAspLysProThrProValAlaLeuTyr 83
23                :||| :||| |||| :|||
24     Db      1532     AGCCCCACATCACGCTGCCCGGCGAGCAGCACTTCGAGGGCGCGCCCTGGCCTTCAG 1591
25
26     QY      84      LeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGly 98
27                ||| :||| ||| ||| ||| ||| |||
28     Db      1592     TTCGAGGCCCGCGACTTCGGCGAAGACCTGTGTCGCGCGCGGCG 1636
29
30     RESULT 28
31     US-08-542-003-1/C
32     ; Sequence 1, Application US/08542003
33     ; Patent No. 5864013
34     ; GENERAL INFORMATION:
35     ; APPLICANT: Goldberg, Edward B.
36     ; TITLE OF INVENTION: MATERIALS FOR THE PRODUCTION OF
37     ; TITLE OF INVENTION: NANOMETER STRUCTURES AND USE THEREOF
38     ; NUMBER OF SEQUENCES: 6
39     ; CORRESPONDENCE ADDRESS:
40     ; ADDRESSEE: Pennle and Edmonds
41     ; STREET: 1155 Avenue of the Americas
42     ; CITY: New York
43     ; STATE: New York
44     ; COUNTRY: US
45     ; ZIP: 10036
46     ; COMPUTER READABLE FORM:
47     ; MEDIUM TYPE: Floppy disk
48     ; COMPUTER: IBM PC compatible
49     ; OPERATING SYSTEM: PC-DOS/MS-DOS
50     ; SOFTWARE: Patent Release #1.0, Version #1.25
51     ; CURRENT APPLICATION DATA:
52     ; APPLICATION NUMBER: US/08/542,003
53     ; FILING DATE: 13-OCT-1995
54     ; CLASSIFICATION: 530
55     ; ATTORNEY/AGENT INFORMATION:
56     ; NAME: Mistrock, S. Leslie
57     ; REGISTRATION NUMBER: 18,872
58     ; REFERENCE/DOCKET NUMBER: 84/71-0005-999
59     ; TELECOMMUNICATION INFORMATION:
60     ; TELEPHONE: (212) 790-9090
61     ; TELEFAX: 212-869-8864
62     ; TELEX: 66441 PENNIE
63     ; INFORMATION FOR SEQ ID NO: 1:
64     ; SEQUENCE CHARACTERISTICS:
65     ; LENGTH: 8855 base pairs
66     ; TYPE: nucleic acid
67     ; STRANDEDNESS: single
68     ; TOPOLOGY: linear
69     ; MOLECULE TYPE: DNA (genomic)
70     ; HYPOTHEetical: NO
71     ; ANTI-SENSE: NO
72     ; ORIGINAL SOURCE:
73     ; ORGANISM: Bacteriophage T4

```

IMMEDIATE SOURCE:
CLONE: TAIL FIBER GENES
US-08-542-003-1

Alignment Scores:

Pred. No.:	271	Length:	8855
Score:	71.50	Matches:	25
Percent Similarity:	46.05%	Conservative:	10
Best Local Similarity:	32.89%	Mismatches:	24
Query Match:	8.43%	Indels:	17
DB:	2	Gaps:	5

US-09-900-575-29_COPY_26_186 (1-161) x US-08-542-003-1 (1-8855)

QY 42 ValLeuSerAsnPhseSerGlyThrValLysTyrSerGlySerSerThrProhProh 61
||||| ||||| :||| :||| ||||| |||||
Db 6757 GTACTACGGAACAGTGCAGAGTAATAGAA-----GCAACAGAAATATCCACCGCACT 6704

QY 62 ThSerGluThrProArgValValTyrAsnSerArgThrAspLysPro-----Trp 78
||||| ||||| :||| :||| ||||| |||||
Db 6703 AGATCAAAATACACAGTTTCTTTGTAT-----GACACGCAAGTTACAGTCTCG 6656

QY 79 ProValAlaLeuTyrLeuThrProVal-----SerSerAlaGlyGly 92
||| ||||| :||| :||| ||||| |||||
Db 6655 CCGAGAACAGATCTATTCGCCCTTACACCGAGCTTGCATATATCACTAGACGGGAA 6596

QY 93 LeuValIleLysAla-----GlySerLeuIleAlaValLeuIleu 106
||||| :||| :||| :||| :||| :|||
Db 6595 TTGTAGTGAATGTACTATATGCAAGGCTATTCAGATATCATCTCTG 6548

RESULT 29
US-08-332-760A-1/C
Sequence 1, Application US/08322760A
Patent No. 5877279

GENERAL INFORMATION:
APPLICANT: Goldberg, Edward B.
TITLE OF INVENTION: MATERIALS FOR THE PRODUCTION OF
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennile and Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: US
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/322,760A
FILING DATE: 13-OCT-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 8471-0003-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: 212-869-8864
TELEX: 66441 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8855 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:

ORGANISM: Bacteriophage T4
IMMEDIATE SOURCE:
CLONE: TAIL FIBER GENES
US-08-332-760A-1

Alignment Scores:

Pred. No.:	271	Length:	8855
Score:	71.50	Matches:	25
Percent Similarity:	46.05%	Conservative:	10
Best Local Similarity:	32.89%	Mismatches:	24
Query Match:	8.43%	Indels:	17
DB:	2	Gaps:	5

US-09-900-575-29_COPY_26_186 (1-161) x US-08-332-760A-1 (1-8855)

QY 42 ValLeuSerAsnPhseSerGlyThrValLysTyrSerGlySerSerThrProhProh 61
||||| ||||| :||| :||| ||||| |||||
Db 6757 GTACTACGGAACAGTGCAGAGTAATAGAA-----GCAACAGAAATATCCACCGCACT 6704

QY 62 ThSerGluThrProArgValValTyrAsnSerArgThrAspLysPro-----Trp 78
||||| ||||| :||| :||| ||||| |||||
Db 6703 AGATCAAAATACACAGTTTCTTTGTAT-----GACACGCAAGTTACAGTCTCG 6656

QY 79 ProValAlaLeuTyrLeuThrProVal-----SerSerAlaGlyGly 92
||| ||||| :||| :||| ||||| |||||
Db 6655 CCGAGAACAGATCTATTCGCCCTTACACCGAGCTTGCATATATCACTAGACGGGAA 6596

QY 93 LeuValIleLysAla-----GlySerLeuIleAlaValLeuIleu 106
||||| :||| :||| :||| :||| :|||
Db 6595 TTGTAGTGAATGTACTATATGCAAGGCTATTCAGATATCATCTCTG 6548

RESULT 30
US-09-236-949-1/C
Sequence 1, Application US/09236949
Patent No. 6437112

GENERAL INFORMATION:
APPLICANT: Goldberg, Edward B.
TITLE OF INVENTION: MATERIALS FOR THE PRODUCTION OF
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennile and Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: US
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/236,949
FILING DATE: 25-Jan-1999
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/542,003
FILING DATE: 13-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 8471-0005-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: 212-869-8864
TELEX: 66441 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8855 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacteriophage T4
IMMEDIATE SOURCE:
CLONE: TAIL FIBER GENES
SEQUENCE DESCRIPTION: SEQ ID NO: 1
US-09-236-949-1

Alignment Scores:
Pred. No.: 271 Length: 8855
Score: 71.50 Matches: 25
Percent Similarity: 46.05% Conservative: 10
Best Local Similarity: 32.89% Mismatches: 24
Query Match: 8.43% Indels: 17
DB: 4 Gaps: 5

US-09-900-575-29_COPY_26_186 (1-161) x US-09-236-949-1 (1-8855)

QY 42 ValLeuSerAsnPhSeSerGlyThrValLysTySerGlySerSerTyProPheProThr 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 6757 GTCACGAGGAACTGTCAGGAGTAATAGAA-----GCAACGATATCCACCGCACT 6704
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 62 ThrSerGluThrProArgValAlaTyAsnSerArgThrAspLysPro-----Trp 78
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 6703 AGATCAATATACCAAGTTTCTTTGTAT-----GACACGCCAGTTACAGTGTG 6656
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 79 ProValAlaLeuTyLeuThrProVal-----SerSerAlaGlyGly 92
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 6655 CCGAGAACAGAACTACTATTCGCCCATACACGCGTCCCATTAATCAGTACGCGGAAA 6596
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 93 LeuValIleLeuSala-----GlySerLeuIleAlaValLeuIleLeu 106
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 6595 TTTCGTAGTGAATGTACTATGCAAGCGTATTACGATATCATCTCTG 6548
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 31
US-08-743-637B-27/c
; Sequence 27, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743.637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435.
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Haemophilus influenzae
US-08-743-637B-27

Alignment Scores:
Pred. No.: 282 Length: 9100
Score: 71.50 Matches: 39
Percent Similarity: 44.53% Conservative: 22
Best Local Similarity: 28.47% Mismatches: 57
Query Match: 8.43% Indels: 19
DB: 2 Gaps: 8

US-09-900-575-29_COPY_26_186 (1-161) x US-08-743-637B-27 (1-9100)

QY 34 GlnArgGlySerAlaTyGlyGlyValLeuSerAsnPhSeSerGlyThrValLysTySer 53
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 884 GAAAAGCCTTTTAT-----TTGATATACGCACTCTGTGTGGGGCTTATCGT 834
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 54 GlySerSerTyProPheProThrThrSerGluThrProArgValAlaTyAsn----- 71
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 833 TTATCCCTCAATGAATCAGAAAAACGGGGAATATAAAGAGTATGGACCGCTGTG 774
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 72 SerArgThrAspLysProTrpProValAlaLeuTyLeuThrProValSer-----Ser 89
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 773 GATCAACAGATTACCGCAAGATATGCTGATGTATAAATGATCAAAAACTCAAT 714
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 90 AlaGlyGlyLeuValIleLys-----AlaGlySer-----LeuIleAlaValLeuIleLeu 106
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 713 TATGCTGCAAGAAAGTACGTATTACGCGGTTCTCCATTAAATGATGCGCGGAATTGTTG 654
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 107 ArgGlnThrAsn-----AsnTyAsnSerAspPheGlnPheValTrp 121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 653 CCGTCACAAGATCGGCCTATGACCGAGCTTCATCTACGATTTTCATCTTCGC 594
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 122 AsnIleTyAlaAsnAspValValValProThrGlyGlyAspValSerAlaArg 141
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 593 ATTAATTTTACGACGACCAAAAATGTC-----ACCGTAAATGTGCAATGCTTCAA 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 142 AspValThrValThrLeuPro-----AspTyArgGlySerValProIle 156
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 539 CAAGTCAAGCAAGCCTTCGCGCAAAAGACGCGGCTTAGTATCATC 489
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 32
US-08-526-840B-27/c
; Sequence 27, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732
FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586,90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 9100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Haemophilus influenzae
US-08-526-840B-27

Alignment Scores:
Pred. No.: 282          Length: 9100
Score: 71.50           Matches: 39
Percent Similarity: 44.53% Conservative: 22
Best Local Similarity: 28.47% Mismatches: 57
Query Match: 8.43%      Indels: 19
                        Gaps: 8

US-09-900-575-29_COPY_26_186 (1-161) x US-08-526-840B-27 (1-9100)
QY 34 GlnArgGlySerAlaTyrglyValIleuSerAsnPheserGlyThrValLysTySer 53
Db 884 GAAAGAGCGCTTTTAT-----TTGGATACGCACCTGTGTGGCGGCGCTTATCCT 834
QY 54 GlySerSerTyrrProPheProThThSerGlyuThrProArgValValTyrrAsn----- 71
Db 833 TTATCTCAATGATGATGAGAAAAACGCGGGAATATTAAGACGATATTGACCGCTGTGG 774
QY 72 SerArgThrAspLysProTrrProValAlaLeuTyrrLeuThrProValSer-----Ser 89
Db 773 GATCAACACAGATTACCCGAAGATATTCCTGCTATGTAATAAGATCAAAAATCTCAAT 714
QY 90 AlaGlyGlyLeuValIleLys-----AlaGlySer---LeuIleAlaValIleuLel 106
Db 713 TATGTGTGCGCAAGACACTGTTATTACGCGGCTTCCTCAATTAATCCGCGGAATTGTG 654
QY 107 ArgGlnThrAsn-----AsnTyrrAsnSerAspAspPheGlnPheValTrrP 121
Db 653 CGCTCAACAAGATGCGCACATATTGACCGAGCTTCATCTTCAAGCATTCATTCCTCCG 594
QY 122 AsnIleTyrrAlaAsnAsnAspValValValProThGlyGlyCysAspValSerAlaArg 141
Db 593 AATAAATTTTACGACGACCAAAAATGTC-----ACCGTAAATGTGCAATGGCTTTCA 540
QY 142 AspValThrValThrLeuPro-----AspTyrrArgGlySerValProIle 156
Db 539 CAAGTCAAAAGCAACGCTTCGCCAAAAGAACGCCGAGCTTACTACTATC 489

RESULT 33
US-08-370-319C-2/C
; Sequence 2, Application US/08370319C
; Patent No. 5856091
; GENERAL INFORMATION:
; APPLICANT: Blichard, Vincent; Van Pel, Aline;
```

```

APPLICANT: Traversari, Catia; W Ifel, Thomas; Coulle, Pierre;
APPLICANT: Boon-Falleur, Thierry; De Plaen, Etienne
TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR PROCESSED TO AT LEAST ONE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Mordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/370,319C
FILING DATE: 10-JANUARY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/272,351
FILING DATE: 8-JULY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/032,978
FILING DATE: 18-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5856091man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5377.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 9421 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
OTHER INFORMATION: Following position there is an
OTHER INFORMATION: unsequenced portion of from 4.7 to 5.3
OTHER INFORMATION: kilobases
US-08-370-319C-2

Alignment Scores:
Pred. No.: 297          Length: 9421
Score: 71.50           Matches: 20
Percent Similarity: 43.94% Conservative: 9
Best Local Similarity: 30.30% Mismatches: 20
Query Match: 8.43%      Indels: 17
                        Gaps: 3

US-09-900-575-29_COPY_26_186 (1-161) x US-08-370-319C-2 (1-9421)
QY 29 AspTyrrValThrLeuGlnArg-----GlySerAlaTyrgly 40
Db 6935 AATATCTGACTCTCCAGATTCAGAGAAAGAACTGGAAGAGCGCCCTGAGACTTGA 6876
QY 41 GlyValIleuSerAsnPheserGlyThrValLysTyrrSerGlySerSerTyrrProPhe--- 59
Db 6875 GGGAT-----TCAGCGGTGGGAACCTTTACCAACCATCAAGGCTGTATCCATTTGCT 6822
QY 60 -----ProThThSerGlyuThrProArgValValTyrrAsnSerArg 73
Db 6821 CTTCTACATACCAACAGCGCATGAGAGATGAGATCCAGGATGTCAGAGATCCG 6762
QY 74 ThrAspLysProTrrPro 79
Db 6761 ATCCAGCGGCGCTGCGCC 6744
```


QY 130 ValValProThrclyGlyCysaspValSerAlaArgaspValThrVal 145
Db 733 TTGGAGGACGACGACAAATTATATTTCGAATTGGAGGTACGGTC 780
RESULT 36
US-09-071-035-263/c
Sequence 263, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071.035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 263:
SEQUENCE CHARACTERISTICS:
LENGTH: 1463 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-263
Alignment Scores:
Pred. No.: 22 Length: 1463
Score: 71.00 Matches: 35
Percent Similarity: 37.32% Conservative: 18
Best Local Similarity: 24.65% Mismatches: 40
Query Match: 8.37% Indels: 49
Gaps: 6
US-09-900-575-29_COPY_26_186 (1-161) x US-09-071-035-263 (1-1463)
QY 19 CysHsAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAla 38
Db 757 TGTTCGAATCAGCTCCATCCTGCATCAGC----- 728
QY 39 TyrGlyValLeuSerAsnPheserGlyThrValIlystYrSerGlySerTyrPro 58
Db 727 -----AATTCGATGCGTACGATTCGATCGCGGTGCGTCCGTCACCG 686
QY 59 PheProThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTr 78
Db 685 ATACACGACACTGAA----- 670
QY 78 pProValAlaLeuThrLeuThrProValSerSerAlaGlyGlyLeuValIleIleVala 98
Db 669 -----TTACCTGTACTCTCTGTACAGTGTGCTGTACACCGTGCATCAGCAG 620
QY 98 ySerLeuIleAlaValLeuIleLeuArgGlnThrAsnSerTyrAsnSerAspPhe 118

Db 619 AACAGGACTTCAGCTCGTTCATT-----TCATTCCCGCTTATCCTTCC 572
QY 118 nPheValTyrPasnIleTyrAlaAsnAspValValThrProThrIleGlyCysAspVa 138
Db 571 AATCAGCTTTACGTA-----TCCCGAGATTGTGCTGCTCCCTGTGT----- 529
QY 138 lSerAlaArgaspValThrValThrLeuProaspTyrArgGlySerAlaProIlePro 158
Db 528 -----AGCGTTACAGTAATAATCACCA---TTGCAATCAGAGGTACTGTACCAAT 482
QY 158 uThr 159
Db 481 GACT 478
RESULT 37
US-09-071-035-257/c
Sequence 257, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071.035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 257:
SEQUENCE CHARACTERISTICS:
LENGTH: 4977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-257
Alignment Scores:
Pred. No.: 134 Length: 4977
Score: 71.00 Matches: 35
Percent Similarity: 37.32% Conservative: 18
Best Local Similarity: 24.65% Mismatches: 40
Query Match: 8.37% Indels: 49
Gaps: 6
US-09-900-575-29_COPY_26_186 (1-161) x US-09-071-035-257 (1-4977)
QY 19 CysHsAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAla 38
Db 2580 TGTTCGAATCAGCTTCATCCTGCATCAGC----- 2551
QY 39 TyrGlyValLeuSerAsnPheserGlyThrValIlystYrSerGlySerTyrPro 58

Db 2550 -----AAATTGATGTAGCATCGATCGATCAGCGGTGCTGTCACCTG 2509
QY 59 Phe-ProThrThrSerGluThrProAlaGlyValTyrAsnSerArgThrAspLysProTr 78
Db 2508 ATACCGACGACCTGAA----- 2493
QY 78 PProValAlaLeuThrProAlaThrProValSerSerAlaGlyGlyLeuValIleLysAlaG1 98
Db 2492 -----TTACCTGTACTCCTGTTACAGTGGTGGCTGTAACCGTCGATCAGCAGG 2443
QY 98 YSerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheG1 118
Db 2442 AACAGGACTTCAGTCGGTGTGACTT-----TCATTTCGCCGTTATCCTTTCC 2395
QY 118 nPheValTrpAsnIleTyrAlaAsnAspValValProThrGlyGlyCysAspVa 138
Db 2394 AATCAGCTTTAACGTA-----TCCCGACGATTGGTCTCCCTTCCGTGGT----- 2352
QY 138 lSerAlaArgAspValThrValThrLeuProAspTyrArgLysValProIleProLe 158
Db 2351 -----AGCGTTACAGTAATAATCACCA---TTCCATCAGAGGTACCTGTACCAAT 2305
QY 158 uThr 159
Db 2304 GACT 2301
RESULT 38
US-09-071-035-261/c
; Sequence 261 Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 261:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-071-035-261
Alignment Scores:
Pred. No.: 134 Length: 4977
Score: 71.00 Matches: 35
Percent Similarity: 37.32% Conservative: 18
Best Local Similarity: 24.65% Mismatches: 40

Query Match: 8.37% Indels: 49
DB: 4 Gaps: 6
US-09-900-575-29_copy_26_186 (1-161) x US-09-071-035-261 (1-4977)
QY 19 CYSNlASnAspTyrProGluThrThrAspTyrValThrLeuGlnArgLysAlaG1 38
Db 2580 TGTTCGAATCAGCTTCCCATCTGCATCAGC----- 2551
QY 39 TyrGlyGlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyrPro 58
Db 2550 -----AATTTCATGATGATACATTGGATCGGCGTCCGTCACTG 2509
QY 59 Phe-ProThrThrSerGluThrProArgValTyrAsnSerArgThrAspLysProTr 78
Db 2508 ATACCGACGACGAA----- 2493
QY 78 PProValAlaLeuThrProAlaThrProValSerSerAlaGlyGlyLeuValIleLysAlaG1 98
Db 2492 -----TTACCTGTACTCCTGTTACAGTGGTGGCTGTAACCGTCGATCAGCAGG 2443
QY 98 YSerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheG1 118
Db 2442 AACAGGACTTCAGTCGGTGTGACTT-----TCATTTCGCCGTTATCCTTTCC 2395
QY 118 nPheValTrpAsnIleTyrAlaAsnAspValValProThrGlyGlyCysAspVa 138
Db 2394 AATCAGCTTTAACGTA-----TCCCGACGATTGGTCTCCCTTCCGTGGT----- 2352
QY 138 lSerAlaArgAspValThrValThrLeuProAspTyrArgLysValProIleProLe 158
Db 2351 -----AGCGTTACAGTAATAATCACCA---TTCCATCAGAGGTACCTGTACCAAT 2305
QY 158 uThr 159
Db 2304 GACT 2301
RESULT 39
US-09-071-035-265/c
; Sequence 265 Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8512
; TELEFAX: (301) 309-8504
; INFORMATION FOR SEQ ID NO: 265:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 4977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-071-035-265

```

Alignment Scores:

File: NO.	134	Length:	4977
Score:	71.00	Matches:	35
Percent Similarity:	37.32%	Conservative:	18
Best Local Similarity:	24.65%	Mismatches:	40
Query Match:	8.37%	Indels:	49
DB:	4	Gaps:	6

200-272-COPY-26-186 (1-161) x US-09-071-035-265 (1-4977)

OY	19	CysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAla	38
Db	2580	TGTTCATCATCAGCTCCCATCTGCATCAC-----	2551
OY	39	TyrgLygIylValLeuSerAsnPheserGlyThrValLysTyrserySergIysSerTyPro	58
Db	2550	-----AATTGGAGTGAGCATTTCCGATACGCCGCTCCTGTACCCTG	2506
OY	59	Phe-ProthrHrserGrluthrProArqValValTyraSnSerArgThrasPlysProTr	78
Db	2508	ATMACCGCAACTGA-----	2493
OY	78	proValAlaleutryLeuthrProvalSerSerAlagIygIyleuValIleLyalagl	98
Db	2492	-----TTACCTGTATCTCTCTGTACAGTGGTGTGCTGTACCGGTGCATCAGCAGG	2443
OY	98	ySerLeuIleAlaValLeuIleLeuArgGlnThrAsnaAntTyraSnSerAspAspphGl	118
Db	2442	AACAGACACTTCAGTCCGTGACTT-----TCATTTCCCGCGTTATCCTTCC	2395
OY	118	nPhelValTrpAsnIleytrylaSnasnsapValValAlProthrrGlyCyAspAspVa	138
Db	2394	AATCACCTGTAAACGTA-----TCCCCAGAGATTGGTCGTTCCCGTGGT-----	2352
OY	138	IserAlaargAspValThrValThrLeuProAspyTyraglySerValProIleProle	158
Db	2351	-----AGCGTTACAGATAAACACACA---TTCCGATCAGAGTAGTACCTGTACCAAT	2305
OY	158	utrh 159	
Db	2304	GACT 2301	

RESULT 40
 US-08-961-527-71
 : Sequence 71, Application US/08961527
 : Patent No. 6420135
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Charles Kunsch
 : TITLE OF INVENTION: streptococcus pneumoniae polynucleotides and Sequences
 : NUMBER OF SEQUENCES: 391
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Human Genome Sciences, Inc.
 : STREET: 9410 Key West Avenue
 : CITY: Rockville
 : STATE: Maryland
 : COUNTRY: USA
 :
 : ZIP: 20850
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 : COMPUTER: HP Vectra 486/73
 : OPERATING SYSTEM: MSDOS version 6.2
 : SOFTWARE: ASCII text
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/961,527
 :
 : FILING DATE:
 :
 : CLASSIFICATION: 424
 :
 : PRIOR APPLICATION DATA:

APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB340P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEO. ID NO: 71:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 32768 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear

US-08-961-527-71

Alignment Scores:	
Pred. No.:	2.49e+03
Score:	70.50
Percent Similarity:	34.40%
Best Local Similarity:	23.40%
Query Match:	6.31%
DB:	4
Length:	32766
Matches:	29
Conservative:	14
Mismatches:	43
Indels:	39
Gaps:	4

US-05-500-5/3-29_COPY_26_186 (1-161) x US-08-961-527-71 (1-32768)

[illegible]

RESULT 41
US-09-295-186-9
Sequence 9, Application US/09295186B
Patent No. 6127137
GENERAL INFORMATION:
APPLICANT: Hasida, Miyoko
APPLICANT: Tsutsunai, No. 61271371KO
APPLICANT: Halkier, Torben
TITLE OF INVENTION: Stringer, Mary Ann
TITLE OF INVENTION: An Acidic Phospholipase, Production, and
FILE REFERENCE: 4953-204-US Methods of Using Thereof (As Amended)
CURRENT FILING DATE: 1999-04-26
CURRENT FILING DATE: US/09/295,186B
PRIOR APPLICATION NUMBER: 1215/96
PRIOR FILING DATE: 1996-10-31
PRIOR APPLICATION NUMBER: PCT/DK97/00490
PRIOR FILING DATE: 1997-10-30

```

; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1870
; TYPE: DNA (genomic)
; ORGANISM: Hypozyma sp. CBS 648.91
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (214)...(1869)
; NAME/KEY: mat_peptide
; LOCATION: (442)...(1869)
US-09-295-186-9

Alignment Scores:
Pred. No.: 42.2 Length: 1870
Score: 70.00 Matches: 33
Percent Similarity: 35.51% Conservative: 16
Best Local Similarity: 23.91% Mismatches: 45
Query Match: 8.25% Indels: 4
Gaps: 7
DB:

US-09-900-575-29_COPY_26_186 (1-161) x US-09-295-186-9 (1-1870)
QY 23 TYRProGluThrIleThrAspTyrValThrLeuGln-----ArgGly 36
    ||| :|||:|||||
DB 1006 TACAGACACGAGTCTCCACGACTTGTGGCTTCGATCGCCGACACAGATTCTCAACGGC 1065
QY 37 SerAlaTyrGlyValLeu-----SerAsnPheserGlyThrValYlsTyr 52
    ||| |||||
DB 1066 TCCATGTACGGCACAAGTTCAGCCGCTGAGTGAGGAGGAGTCAAGAAATACGTCACAAATTC 1125
QY 53 SerGlySerSerTyrProPheProThrThrSerGlyThrProArgValYlsAsnSer 72
    :|||:|||||
DB 1126 ACCGATGCTCCATCGCTGCCCTCCCATC-----ATTATGCCGACGAG 1167
QY 73 ArgThrAspIysProThrProValAlaLeuTyrLeuThrProValSerSerAlaGly 92
    |||
DB 1168 CGC----- 1170
QY 93 LeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnSerTyr 112
    :|||:|||||
DB 1171 -----GAGCGCGGCGAGCTCATC-----ATCCCGCGCAACACACCATCTGG 1212
QY 113 AsnSerAspPheGlnPhe---ValTrpAsnIleTyrAlaAsnAsnAspValVal 131
    :|||:|||||
DB 1213 GAGTTCAACCGCAGCGAGTTCGCTTGGAAC-----CCCAATGTTCCGCTTTCATC 1266
QY 132 ProThrGlyGlyCysAspValSerAlaArgAspValThrValThrLeuProAsp 149
    ||| |||
DB 1267 CCCATCGAGATCCCTCGGCTCGAGTCTGGACACGCGACGCTCTGCCCGAC 1320

RESULT 42
US-08-976-259-85/c
; Sequence 85, Application US/08976259
; Patent No. 6316609
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Choi, Gil H.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: Nucleotide sequence of Escherichia coli
; Patent No. 6316609
; NUMBER OF SEQUENCES: 142
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
```

```

; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,259
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,626 AND US 60/061,953
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CBM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-976-259-85

Alignment Scores:
Pred. No.: 521 Length: 9319
Score: 69.50 Matches: 32
Percent Similarity: 40.00% Conservative: 28
Best Local Similarity: 21.33% Mismatches: 71
Query Match: 8.20% Indels: 19
Gaps: 4
DB:

US-09-900-575-29_COPY_26_186 (1-161) x US-08-976-259-85 (1-9319)
QY 26 ThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyValLeuSerAsn 45
    |||:|||||
DB 7073 ACCGTGAGTATACAGGACGAGATACAGGAGTGT-----GGCGGCTTCCCTGGAAC 7023
QY 46 PheSerGlyThrValYlsTyrSerGlySerSerTyrProPheProThrThrSerGly 65
    :|||
DB 7022 GCCACTGACCGCTGCGCAACAGCAGCGCAAAATCTCTCCGCACTAACCTCAGCGTACG 6963
QY 66 ProArgValIleTyrAsnSerArgThrAspIysProThrProValAlaLeuTyrLeuThr 85
    :|||:|||||
DB 6962 GCGCAGGTCTCTGCGCAACACGCGGAGCTGTGATACAGGCTGCCACCTCTCTCTGAT 6903
QY 86 ProValSerSerAla---GlyGlyLeuValIleLysAlaGlySerLeu---IleAlaVal 103
    |||:|||||
DB 6902 GTGGTAATACTGTCAACGCGGAGCGTACTTGCACCGGCAAGGCCGACGTTAAAGA 6843
QY 104 LeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheValTrpAsnIle 123
    ||| |||
DB 6842 ACCACGCTGAATATATACCGGTACGCTTCAGGCTGGCGACCTGTGTGATTAACACACA 6783
QY 124 TyrAlaAsnAsnAspValValProThrGlyCysAspVal----- 138
    :|||:|||||
DB 6782 TTCAGCAACAGCGGATACCTGCTGGAACTTCGGGCTTGGGCTGAAGGAGATTCAC 6723
QY 139 -----SerAlaArgAspValThrValThrLeuProAsp 149
    ||| |||
DB 6722 CTCGAATATGTAACAGGCGGCTGTACAGTGCAAGCAACCTGCTGTACCGCTCAGGAC 6663
QY 150 TyrArgGlySerValProIleProLeuThr 159
    |||
DB 6662 TTCAGTGTACAGGCGAGGTGTGGCGCAC 6633

RESULT 43
US-08-976-259-14
; Sequence 14, Application US/08976259
; Patent No. 6316609
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Choi, Gil H.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: Nucleotide sequence of Escherichia coli
```

```
Patent No. 6316609
NUMBER OF SEQUENCES: 142
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Ave, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,259
FILING DATE: Herewith
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/031,626 AND US 60/061,953
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488, 0740002/EKS/CBM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
FAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 22671 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-976-259-14

Alignment Scores:
Pred. No.: 1,93e+03 Length: 22671
Score: 69.50 Matches: 32
Percent Similarity: 40.00% Conservative: 28
Best Local Similarity: 21.33% Mismatches: 71
Query Match: 8.20% Indels: 19
DB: 4 Gaps: 4

US-09-900-575-29_COPY_26_186 (1-161) x US-08-976-259-14 (1-22671)
Qy 26 ThrlethraspyrvalThleuGlnArglySerAlatryglyValleuSerasn 45
Db 16130 ACCCTGAGTACCGGCGACGATACAGGTGTT-----GGCGGGTTTCCCTGAC 16180
Qy 46 PheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThr 65
Db 16181 GCCACTGACCGTCTGCAGAACGACGCAAAATCCTCCGCGAGTAACTCAGCGTGACG 16240
Qy 66 ProAlaValAlaTyrAsnSerArGThrAspLysProTyrProValAlaLeuTyrLeuThr 85
Db 16241 GCCCAGAGTCTGCGCAACACGCGACGCGACTGTCACAGCTCCACCTGCTGGTGAAT 16300
Qy 86 ProValSerSerAla----GlyGlyLeuValIleLysAlaGlySerLeu---IleAlaVal 103
Db 16301 GTGCTGAATACCTGCAACGCGGACGCGTACTGTCACCGCGCTCCAGCTTAAGA 16360
Qy 104 LeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheValItrpAsnIle 123
Db 16361 ACCACGCTGAATATACCGGTACGCTTCAGGGTCCGACCTGCTGTAATTACACACA 16420
Qy 124 TyrAlaAsnAsnAspValAlaValProThnrglyGlyCysAspVal----- 138
Db 16421 TTCAGAACAGCGGTACCTGCTGGGAACTCCGGGCTTGGGTGAAGGCGAGTTCAC 16480
Qy 139 -----SerAlaArgAspValThValThreProAsp 149
Db 16481 CTGCAGAAATGCTACAGCGGCTGTAAGTGAAGCAACCTGCTGTCAGCGCTCAGAC 16540
```

```
Qy 150 TyrArglySerValProIleProLeuThr 159
Db 16541 TTCAGTGTGTCAGGGCGACGCTGTCGCCACC 16570
RESULT 44
US-08-712-072C-1
Sequence 1, Application US/08712072C
Patent No. 5925541
GENERAL INFORMATION:
APPLICANT: Jack Goldstein, Alex Zhu and Lin Leng
TITLE OF INVENTION: ENDO-BETA-GALACTOSIDASE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amster, Rothstein & Ebenstein
STREET: 90 Park Avenue
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/712,072C
FILING DATE: 11-SEP-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bogosian, Elizabeth A.
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/97
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
FAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1408 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-712-072C-1

Alignment Scores:
Pred. No.: 37.1 Length: 1408
Score: 69.00 Matches: 48
Percent Similarity: 33.49% Conservative: 23
Best Local Similarity: 22.64% Mismatches: 67
Query Match: 8.14% Indels: 74
DB: 2 Gaps: 11

US-09-900-575-29_COPY_26_186 (1-161) x US-08-712-072C-1 (1-1408)
Qy 12 AspleuSerThrGlnIlePheCysHisAsnAsp----- 22
Db 711 AACCTATATACGATGATCTGGAGTCCGACACATTCGATTCTACGTAACATTCAATTG 770
Qy 23 -----TyrProGluThrIleThr 28
Db 771 CAGTATACCTACGCAAGAGTTTCGGTGGGGGACACAGCAATGGCCATTGACGTTCC 830
Qy 29 AspTyrValThrLeuGlnArg-----GlySerAlaTyrGlyValLeuSerAsn--- 45
Db 831 TTTATCTGATTCTTAATCAAGCGCGGTGAGCGGATGCCAGGCGGAGATGCACAAATGCT 890
```

```

QY 46 -----PheSerGlyThrValLysTyr----- 52
Db 891 GACTTGCCCTTACGTAGCAGCGATGAGTGGCTGATATAGTGCCTTATTTAGT 950
QY 53 -----SerGlySerSerTyrProPheProThrThrSerGlyThrProArg 67
Db 951 AATGCCGATTCGAAAGGGGTGCATCTATCCATGACACATGGCGGATGATCAGC 1010
QY 68 ValValTyrAsnSer-----ArgThrAspLysProTyr 78
Db 1011 GTTGTTCACCGAGCCCGGACAGCAACAAATGATCCCGCAACGGCGGAGACACA 1070
QY 79 ProValAlaLeuTyrLeuThrProValSer-----SerAlaGlyLeu 93
Db 1071 TCCATTGACAACTACTGACCGGTTTACGCCAATAGCACTTATCGGTGGTGCTAC 1130
QY 94 -----ValIleLysAlaGlySerLeuIleAlaValIleLeuArgGlnThrAsn 111
Db 1131 GCCAAGTGTGACGTCGACGTCGCAATCAGTATGTTGGTGC-----AAAAT 1178
QY 112 TyrAsnSerAsp-----AspPheGlnPheValTyrAsnIleTyrAlaAsnAspVal 129
Db 1179 TATGGGGGAACTCGCGTGCATGACTAGTAGACAGCACTACTCCAAATATTCGCTA 1238
QY 130 ValValProThrGlyGlyCysAspValSerAlaArgAspValThrValThrLeuProAsp 149
Db 1239 ACTTTTACAACTGGA-----GCCAATAATCACTACTGCTACGCTCTATTTC 1283
QY 150 TyrArgGlySerValProIleProLeuThrValTyr 161
Db 1284 TATAAA-----CCCTTGAGCGGTACAGTGTAT 1310

```

RESULT 45

```

US-09-620-412C-308
; Sequence 308, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469c7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 308
; LENGTH: 1860
; TYPE: DNA
; ORGANISM: Chlamydia trachomatis
US-09-620-412C-308

```

```

Alignment Scores:
Pred. No.: 55.9 Length: 1860
Score: 69.00 Matches: 35
Percent Similarity: 39.87% Conservative: 26
Best Local Similarity: 22.88% Mismatches: 70
Query Match: 8.14% Indels: 23
DB: Gaps: 7

```

US-09-900-575-29_COPY_26_186 (1-161) x US-09-620-412C-308 (1-1860)

```

QY 8 AsnLeuValValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGlu----- 25
Db 1130 AACATAGTGTGTC--GTGGAGGAGCAATATGCGGTGACTTCCGACGATCCAGAAATTCTT 1186
QY 26 ---ThrIleThrAspTyrValThrLeuGlnArgGlySerAla---TyrGlyGlyValLeu 43
Db 1187 TTCTTAATAACCAACA-ATTACTTTCGAAAGCAATAGCGCTGTGCATGAGAGTGTATC 1245
QY 44 SerAsnPheSerGlyThrValLysTyrSerGlySerTyrProPhePro----- 60
Db 1246 TACATAAGAAATGCGCTTGTGAGTTCTTAGAAATGCGAGACCTTGGCTTAAAGAG 1305

```

```

QY 61 ---ThrThrSerGlnThrProArgValValTyrAsnSerArg-----ThrAspLys 76
Db 1306 AACACAACAATAGCTAACGGGGAGCTATATACACAAGTAATTTAAAGCGAATCAACA 1365
QY 77 ProTyrProValAlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLys 96
Db 1366 ACATCCCGCATTTCTATTCTCAAAATCATGGGAATPAAGAAAGCGGGA----- 1413
QY 97 AlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAsp 116
Db 1414 -----GCGATTTACGGCAATATGTGAACCTTAGAACAGATCAAGTACT 1458
QY 117 PheGlnPheValTyrAsnIleTyrAlaAsnAspValValValProThrGlyLeu 136
Db 1459 ATTGGCTTTGAAAAAATTAACCGCTAAAGAGCGGTGAGCCATCACTCTCTCAATGC 1518
QY 137 AspValSerAlaArgAspValThrValThrLeuProAsp 149
Db 1519 TCAATTACTGCTCATATAT---ACCATCACTTTTCCGAT 1554

```

Search completed: November 28, 2002, 20:44:20
Job time : 778 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 28, 2002, 19:59:16 ; Search time 53 seconds

(without alignments)
1169.879 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186

Perfect score: 848
Sequence: 1 FVVNVGNLVDLSTQIFCH.....DVTYLPDYGVSPIPLTV 161

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 341543 seqs, 19257720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-O=/cgn2_1/USPto.spool/US09050575/runtat_22112002_130710_4669/app_query.fasta_1.327
-DB=Published.Applications.NA -OFTM=fastap -SUFFIX=rmph -MINMATCH=0.1
-LOOCL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100
-THR.MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFM=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09050575 -ECGN_1_1_21=/runtat_22112002_130710_4669
-NCPU=6 -ICPU=3 -NO.XLPRY -NO.MMAP -LARGEDBURY -NEG.SCORES=0 -WAIT -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	848	100.0	837	10	US-09-900-575-7
2	841	99.2	837	10	US-09-900-575-17
3	834	98.3	837	10	US-09-900-575-6
4	834	98.3	837	10	US-09-900-575-21

5	834	98.3	840	10	US-09-900-575-10	Sequence 10, Appl
6	834	98.3	840	10	US-09-900-575-12	Sequence 12, Appl
7	834	98.3	903	10	US-09-912-020-20	Sequence 207, Appl
8	834	98.3	903	10	US-09-900-575-48	Sequence 48, Appl
9	834	98.3	7416	10	US-09-900-575-46	Sequence 46, Appl
10	834	98.0	837	10	US-09-900-575-54	Sequence 54, Appl
11	830	97.9	837	10	US-09-900-575-54	Sequence 3, Appl
12	826	97.4	837	10	US-09-900-575-15	Sequence 15, Appl
13	825	97.3	837	10	US-09-900-575-15	Sequence 20, Appl
14	824	97.2	837	10	US-09-900-575-5	Sequence 18, Appl
15	822	96.9	837	10	US-09-900-575-20	Sequence 16, Appl
16	822	96.9	837	10	US-09-900-575-16	Sequence 8, Appl
17	820	96.7	837	10	US-09-900-575-18	Sequence 11, Appl
18	819	96.6	837	10	US-09-900-575-11	Sequence 22, Appl
19	819	96.6	837	10	US-09-900-575-22	Sequence 4, Appl
20	819	96.6	840	10	US-09-900-575-4	Sequence 16, Appl
21	818	96.5	837	10	US-09-900-575-9	Sequence 13, Appl
22	817	96.3	837	10	US-09-900-575-1	Sequence 9, Appl
23	816	96.2	837	10	US-09-900-575-13	Sequence 1, Appl
24	816	96.2	837	10	US-09-900-575-13	Sequence 1, Appl
25	812	95.8	837	10	US-09-900-575-14	Sequence 14, Appl
26	812	95.8	837	10	US-09-900-575-2	Sequence 2, Appl
27	77.5	9.1	1777	10	US-09-910-575-19	Sequence 19, Appl
28	76	9.0	1904	10	US-09-745-763-145	Sequence 15, Appl
29	76	9.0	1977	12	US-10-032-382-79	Sequence 14, Appl
30	76	9.0	1977	12	US-10-032-382-79	Sequence 15, Appl
31	75.5	8.9	1077	10	US-09-815-242-503	Sequence 79, Appl
32	75.5	8.9	1346	10	US-09-815-242-4841	Sequence 8841, Ap
33	75.5	8.9	1356	10	US-09-815-242-4751	Sequence 4751, Ap
34	74	8.7	17056	10	US-09-815-242-9029	Sequence 9029, Ap
35	73.5	8.7	17397	10	US-09-783-066-8	Sequence 8, Appl
36	73.5	8.7	17397	10	US-09-783-066-8	Sequence 8, Appl
37	73.5	8.7	19334	10	US-09-764-869-1943	Sequence 1943, Ap
38	73.5	8.7	19334	10	US-09-764-869-1943	Sequence 1944, Ap
39	73.5	8.7	19345	9	US-09-764-869-1944	Sequence 2387, Ap
40	72.5	8.5	6410	10	US-09-922-217-1094	Sequence 1094, Ap
41	72.5	8.5	17335	10	US-09-922-217-1094	Sequence 1280, Ap
42	72.5	8.5	19882	10	US-09-764-847-1280	Sequence 1281, Ap
43	71.5	8.4	882	10	US-09-764-847-1281	Sequence 149, Ap
44	71.5	8.4	8220	10	US-09-925-289-149	Sequence 3, Appl
45	71.5	8.4	9100	10	US-09-797-808-3	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-09-900-575-7
Sequence 7, Application US/09050575
Patent No. US20020150587A1
GENERAL INFORMATION:
APPLICANT: Langemann, Solomon
APPLICANT: Revel, Andrew
APPLICANT: Auguste, Christine
APPLICANT: Burteln, Jeanne
TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
FILE REFERENCE: 469201-549
CURRENT APPLICATION NUMBER: US/09/900, 575
PRIOR FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: US/60/216, 750
NUMBER OF SEQ ID NOS: 64
SOFTWARE: PatentIn version 3.0
SEQ ID NO 7
LENGTH: 837
TYPE: DNA
ORGANISM: E. coli
US-09-900-575-7

Alignment Scores:
Pred. No.: 1.09e-104
Score: 848.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%

Length: 837
Matches: 161
Conservative: 0
Mismatches: 0

Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-7 (1-837)

```

OY 1 ProvalValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
   |||||||
Db 76 CCCGCGTGAATGTGGGCAAAACCTGGCTGGATCTTTCACACCAATCTTTGGCCAT 135
OY 21 AsnAspTyrProGlnThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
   |||||||
Db 136 AACGATTATCCGGAACCATTCACACTATGTCACACTGCACACGAGCTCGCTTATGGC 195
OY 41 GlyValLeuSerAsnPhseSerGlyThrValLysTyrSerGlySerTyrProPhePro 60
   |||||||
Db 196 GCGCGTTATCTAATTTTCCGGGACCTAAATATAGTACAGTATGCTATCCATTTCCT 255
OY 61 ThrThrSerGlnThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
   |||||||
Db 256 ACCACACGCAAAACGCCGCGCTGTTATATTCGAGAACGAGATAACCGCTGGCGGTG 315
OY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
   |||||||
Db 316 GCGCTTATTTGACGCTGTGACAGCTGCGGGGCTTGATTAAGCTGGCTCATTTA 375
OY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
   |||||||
Db 376 ATGGCGTCTTATTTTGGACAGCAACACTATACAGGAGATTTCAGATTGTG 435
OY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
   |||||||
Db 436 TGGAAATTTACGCCAATATGATGTGTGCTACTAGCGGCTGGCATTTCTGCT 495
OY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
   |||||||
Db 496 CGTGATGTACCGTTACTGTGCGGACTACCGGTTCACTGCAATTCCTTACCGTT 555
OY 161 Tyr 161
   |||
Db 556 TAT 558

```

RESULT 2

```

; US-09-900-575-17
; Sequence 17, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burielin, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-17

```

Alignment Scores:

Pred. No.: 9.52e-104
Score: 841.00
Percent Similarity: 99.38%
Best Local Similarity: 98.76%
Query Match: 99.17%

Length: 837
Matches: 159
Conservative: 1
Mismatches: 1
Indels: 0
Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-17 (1-837)

```

OY 1 ProvalValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
   |||||||
Db 76 CCCGCGTGAATGTGGGCAAAACCTGGCTGGATCTTTCACACCAATCTTTGGCCAT 135
OY 21 AsnAspTyrProGlnThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
   |||||||
Db 136 AACGATTATCCGGAACCATTCACACTATGTCACACTGCACACGAGCTCGCTTATGGC 195
OY 41 GlyValLeuSerAsnPhseSerGlyThrValLysTyrSerGlySerTyrProPhePro 60
   |||||||
Db 196 GCGCGTTATCTAATTTTCCGGGACCTAAATATAGTACAGTATGCTATCCATTTCCT 255
OY 61 ThrThrSerGlnThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
   |||||||
Db 256 ACCACACGCAAAACGCCGCGCTGTTATATTCGAGAACGAGATAACCGCTGGCGGTG 315
OY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
   |||||||
Db 316 GCGCTTATTTGACGCTGTGACAGCTGCGGGGCTTGATTAAGCTGGCTCATTTA 375
OY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
   |||||||
Db 376 ATGGCGTCTTATTTTGGACAGCAACACTATACAGGAGATTTCAGATTGTG 435
OY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
   |||||||
Db 436 TGGAAATTTACGCCAATATGATGTGTGCTACTAGCGGCTGGCATTTCTGCT 495
OY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
   |||||||
Db 496 CGTGATGTACCGTTACTGTGCGGACTACCGGTTCACTGCAATTCCTTACCGTT 555
OY 161 Tyr 161
   |||
Db 556 TAT 558

```

RESULT 3

```

; US-09-900-575-6
; Sequence 6, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burielin, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-6

```

Alignment Scores:

Pred. No.: 8.3e-103
Score: 834.00
Percent Similarity: 98.76%
Best Local Similarity: 98.14%
Query Match: 98.35%

Length: 837
Matches: 158
Conservative: 1
Mismatches: 2
Indels: 0
Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-6 (1-837)

```

OY 1 ProvalValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
   |||||||
Db 76 CCCGCGTGAATGTGGGCAAAACCTGGCTGGATCTTTCACACCAATCTTTGGCCAT 135

```


Oy	21	ASNAAPTYRPROGLUTHTRIETHTHRSPHYRVALTHRLAUGLNARGLYSERLATHRYGLY	40
Db	136	AACGATTTACCCCGGAAACCATTAACGATTAATGTCACACTGCACAACGAGGCTCGGCTTAAGGC	195
Oy	41	GIYVALLLEUSERASPHESESGLYTHRVALLYSTRYSERGLYSERSETTYRPROPHPRO	60
Db	196	GGCGGTATATCAATTTTTCGGGACCTTAATAATATGTCGCAATACCTATACCAATTCGCG	255
Oy	61	THRTTISRGLUTHPROARGVALVALTYRASNSEARGLTHRRASPLYSPTOTRPPROVAL	80
Db	256	ACCACCACTGGAACCCGCCGGTGTGTTATAATTCGAGAACGATTAAGCCGTGCCGCTG	315
Oy	81	ALALEUTYRLEUTHTHPPROVALSERLASERLAELIYLEUVALILEYLSAIGLYSERLEU	100
Db	316	GCGCTTATTTTGACGCCCTGTGAGAGTGGCGGGCGGGTGGTGATTAACCTGGCTCATTA	375
Oy	101	ILEALVALLEUILELEARGGLNTHRASNAHTYRASNSEARSPASPHGILNPHVAL	120
Db	376	ATTCCCGCTTATTTTGGCAGACCAACACATATAACGCGATGATTTCCAGTTGGTG	435
Oy	121	TRPASNIETRYRALAASNAASNPVALVALALPROTHTRIYGLYCYSAPVALSERALA	140
Db	436	TGGAAATTTTACGGCAATTAATGATGTGTGTGGCCCTACCTGGCGGTGCGATGTTTCGT	495
Oy	141	ARGASPYALTHRVALTHRLNPPOASPTTYRARGLYSERVALPROLLEPROLEUTHRVAL	160
Db	496	CGTGATGCAACCGTTACTCTGCCGAGCTACCCCTGGTTCAGTGCCGATGCCGATCCTCTTACCGTT	555
Oy	161	TYR 161	
Db	556	TAT 558	

```

US-09-900-575-21
: Sequence 21, Application US/09900575
: Patent No. US20020150587A1
: GENERAL INFORMATION:
: APPLICANT: Langermann, Solomon
: APPLICANT: Revel, Andrew
: APPLICANT: Auguste, Christine
: APPLICANT: Builein, Jeanne
: TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
: FILE REFERENCE: 469201-549
: CURRENT APPLICATION NUMBER: US/09/900,575
: CURRENT FILING DATE: 2001-07-06
: PRIOR APPLICATION NUMBER: US/60/216,750
: NUMBER OF SEQ ID NOS: 64
: SOFTWARE: Patentin version 3.0
: SEQ ID NO 21
: LENGTH: 837
: TYPE: DNA
: ORGANISM: E. coli
US-09-900-575-21

Alignment Scores:
Pred. No.:      8.3e-103
Score:          834.00
Percent Similarity: 98.768
Best Local Similarity: 98.144
Query Match:    98.358
DB:             10
                Gaps: 0
                Matches: 837
                Conservative: 158
                Mismatches: 2
                Indels: 0

```

US-09-900-575-25-COPY₂₆-186 (1-161) x US-09-900-575-21 (1-837)

OY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20

Db 76 CCGCGTGTAAATGCTGGGCAAAACCTGCTGGATCTTCCACCGAATCTTTGGCAT 133

OY 21 AsnAspTyr-ProGlnThrIleThrAspTyrValThrLeuGlnArgLysSerIleArgLys 40

Db 136 AAGGATTATTCGGAACCATATTACGACTATGTCACACGAGAGGCTGCGATTATGCG 195

Qy	41	GlyValIeuSerInPheSerIythrValLysTySerIySerSerTyTyrProPhePro	60
Db	196	GGCGTGTATCTAATATTTTTCCGGAGCCGTAATAATATAGGGAGTATGCTATTCATTCT	255
Qy	61	ThrThrSerGlnThrProAlaGlyValLysTyraaSerAlaGlyThrAspLysProTyrProVal	80
Db	256	ACCAACGACGAAACGCCGGCGCGTGTATTAATTCAGACAGCATMAACCCGTGGCCGGTG	315
Qy	81	AlaLeuTyrIleuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu	100
Db	316	GGCGTTATTTGACGGCGCTGAGACAGTGGCGGGCGGGCGTTCATTAAGCGGCTCATTTA	375
Qy	101	IleAlaValLeuIleLeuArgLInThrAsnAsnTyraaSerAspAspPheGlnPheVal	120
Db	376	ATTGGCGGTGCTAATTTTTGGACAGACCAACAACATTAACAGCGATGATTCGACTTTGTC	435
Qy	121	TyrAsnIleTyrAlaAsnAsnAspValValProThrArgLysGlyCysAspValSerAla	140
Db	436	TGGAAATATTTACGCCAATATATATGTGGTGGCTTACGTGGCGCTCGAGTTCCTCGCT	495
Qy	141	ArgAspValThrValThrLeuProAspTyrArgLysSerValProIleProLeuThrVal	160
Db	496	CGGATGTCACCGTTACTCTGCGGAGATACCCGTGGTTGAGTCCCAATTCCTTTACCGTT	555
Qy	161	Tyr 161	
Db	556	TAT 558	

```

RESULT 5
US-09-900-575-10
; Sequence 10, Application US/09000575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Butlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 840
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-10

```

Pred. NO.:	8.35e-103	Length:	840
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	10	Gaps:	0

US-09-900-5/5-29_COPY_26_186 (1-161) x US-09-900-575-10 (1-840)

QY	1	ProValValAlaValG1G1G1AsnLeuValValAspLeuSerThcGlnIlePheCysHis	20
Db	76	CCCGCGGTGAATGGGGCAAAACCTGCTGTGCATCTTTCACGCCAAATCTTTGGCAT	135
QY	21	AsnAspTyrProGlnThIleThrAspTyrValThLeuGlnArgGlySerAlaTyrGly	40
Db	136	AACGATTACCCGGAAACCATACGATTATGTACACTCCAAACGAGGCTCGCGTTATAGGC	169
QY	41	GlyValLeuSerAsnPheSerGlyThhValLysTyrSerGlySerSerTyrProPhePro	60
Db	196	GGGGGTGTTATCTCAATTTTTCCGGGACCGTAAATATAGGAGCAATGTTATCTATTTTCC	355

QY	81	AlAlaLeuTyrIleuThrProValSerSerAlaGlyValIleLeuAlaGlySerIleu	100
		GGCGTTTATTTTGGACGCTGTGGACCACTGGGGGGGGGGCGCATTTAAAGCTGGCTCATTA	15433
Db	1484		
QY	101	IlAlaValLeuIleLeuArgGlnThrAsnTyrAsnSerAspAspPheGlnPheVal	120
Db	1544	ATTGCGCTGGCTATTTTGGCAGCAGACCAACACTATACGCGGATGATTTCCAGTTTG	16030
QY	121	TrpAsnIleTyrAlaAsnAspValValValProThrGlyGlyCysAspValSerAla	140
Db	1604	TGGAAATTTTACGCCAATTAATGATGGGAGGAGGCTACTAGCGCGCGCTGCATGTTTCTGCT	1665
QY	141	ArgAspValThrValThrIleuProAspTyrArgGlySerValProIleProLeuThrVal	160
Db	1664	CGGTATGTCAACCGTTACTCTGCGGAGCACTACCTGGTTCAAGTCCCAATTCCTTTACCGTT	1723
QY	161	Tyr 161	
Db	1724	TAT 1726	

```

RESULT 10
US-09-900-575-54
; Sequence 54, Application US/09900575
; Patent NO. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langemann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Bulein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 54
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-54

```

Alignment Scores:	
pred. No.:	2,1e-102
Score:	831.00
Percent Similarity:	98.75%
Best Local Similarity:	97.52%
Query Match:	98.00%
DB:	10
Length:	837
Matches:	157
Conservative:	2
Mismatches:	0
Indels:	0
Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-54 (1-837)

QY	1	ProValValAsnValGlyGlnAsnLeuValValAspSerSerThrGlnIlePheGlySHis	20
Db	76	CCCGTCGGAGATGGGGGCAAAACCTGGCTGGATCTTTCGACCAACTTTTGCCAT	135
QY	21	AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly	40
Db	136	AACGATTTATCCGAACACCATTAACAGATATGTCACCTGCACAGACGCTCGGCTTATGGC	195
QY	41	GlyValIleuSerAspPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro	60
Db	196	GGCGGTATCTAAATTTTCCGGGACCGTAAATAATATAGGCACTAGCTATCCATTCCCT	255
QY	61	ThrThrSerGluThrProArgValValAlnTyrAsnSerArgThrAspLysProTTPProVal	80
Db	256	ACCACACAGGAACCCCGCGTGTATTATTCGAGAACGGAATAGCCGTGACCGGTG	315
QY	81	AlaLeuTyrLeuThrProValSerSerIleGlyGlyLeuAlaIleValIleGlySerLeu	100
Db	316	GGCGTTATTATACCCCTGTGACAGTCCGGGGGGGGGCGGATTAAGCTGGCTCATTTA	375

QY	101	11	lealavalleu11eleaavg1inthasnas1tyr1asnserspsap1p1eg1in1p1eval	120
Db	376	ATTGGCGGCTTATT	TTCGCACACACCACACATATACACCGATGATTTCACGTTTGT	435
QY	121	TrpAsn11eTyrrAlaasnasasPvalVal1ProThrglyglycysasPval1SerAla	140	
Db	436	TGGATATTTCACCAATAAATGATGTGTGTGCTCTACGGGGCTCGATGTTTCGCT	495	
QY	141	Argaspval1Thrval1ThLeuproasp1tyr1Arg1glyserVal1Pro1LeuPro1Leu1Thval	160	
Db	496	CGTGATGTCACCGTTACTCTCCGCGACTACCCGTGGTTCAGTGCACAAATTCCTTTAACGGTT	555	
QY	161	Tyr 161		
Db	556	TAT 558		

```

RESULT 11
US-09-900-575-3
? Sequence 3. Application US/09005575
? Patent No. US20020150587A1
? GENERAL INFORMATION:
? APPLICANT: Langemann, Solomon
? APPLICANT: Revel, Andrew
? APPLICANT: Auguste, Christine
? APPLICANT: Burlin, Jeanne
? TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
? FILE REFERENCE: 469201-549
? CURRENT APPLICATION NUMBER: US/09/900,575
? CURRENT FILING DATE: 2001-07-06
? PRIOR APPLICATION NUMBER: US/60/216,750
? PRIOR FILING DATE: 2000-07-07
? NUMBER OF SEQ ID NOS: 64
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 3
? LENGTH: 857
? TYPE: DNA
? ORGANISM: E. coli
US-09-900-575-3

```

Alignment Scores:	
Pred. No.:	2,86e-102
Score:	830.00
Percent Similarity:	98.14%
Best Local Similarity:	97.52%
Query Match:	97.68%
DB:	10
	Gaps:
	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-3 (1-837)

QY	1	ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis	20
Db	76	CCGCGCGTGAATGATGGGGCAAAACCTGGTCGTAGTCTTTTCGACGCAAACTTTGGCAT	135
QY	21	AsnAspTryProGluThrThrLeuAspPheValThrLeuGlnAcgLYserAlaTryGly	40
Db	136	AACGATTATCCGGAAACCATTCACACATATGTACACCTGCACACAGGCTGGGTTATGGC	195
QY	41	GlyValLeuSerAsnPheSerGlyThrValLYsTrySerGlySerSerTryProPhePro	60
Db	196	GCGGTATTATCAATTTTCCGGGACCGTAAATATATGACGACTAGCTATCCATTTCCG	255
QY	61	ThrThrSerGluThrProArgValValLYrAsnSerAlaGlnAspLYsProThrProVal	80
Db	256	ACCACACAGCAAAACCGCGCGGCTTTATATAATTCGAACAGGATAAAGCCGTGCCGCTG	315
QY	81	AlaLeuThrLeuThrProValSerIleAsnGlyGlyLeuValIleLYsAlaGlySerLeu	100
Db	316	GCGCTTATTATTCAGCGCTTGAGCAGTGGGGGGGGGCGCATTTAAAGCTGGCTCATTA	375
QY	101	IleAlaValLeuIleLeuArgGlnThrAsnAsnTryAsnSerAspAspPheGlnPheVal	120
Db	376	ATTGGCGCTGTTATTTTGGCAGACAGCAACACTTAACACAGATGATTTCCATTTGTG	435

QY	121	TRPSNIIEETRYALAAANASPAVALVALAIPROHRTGLYGLYCYSASPVALSERALA	140
Db	436	TGGAAATPTTTTACGCCAATAATGATGCTGCTGCTCTACTGGGGCTCGAATGTTTCTGCT	499
QY	141	ARGASPAVALTRVALTPHLEUPROASPTYVALARGLISERVALPROLIEPPOLEUTHVAL	160
Db	496	CGTGATGTCACACCGTTACTCTCCGCGAATAACCTGTTCAGTGGCAAAATCTCTTTAACCGTT	555

RESULT 12

```

: Sequence 15, Application US/09900575
: Patent No. US20020150587A1
: GENERAL INFORMATION:
: APPLICANT: Langermann, Solomon
: APPLICANT: Revel, Andrew
: APPLICANT: Auguste, Christine
: APPLICANT: Burstein, Jeanne
: TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
: FILE REFERENCE: 469201-549
: CURRENT APPLICATION NUMBER: US/09/900,575
: CURRENT FILING DATE: 2001-07-06
: PRIOR APPLICATION NUMBER: US/60/216,750
: PRIOR FILING DATE: 2000-07-07
: NUMBER OF SEQ ID NOS: 64
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 15
: LENGTH: 837
: TYPE: DNA
: ORGANISM: E. coli
: IS-09-900-575-15

```

Alignment Scores:	
Pred. No.:	9,866-102
Score:	826.00
Percent Similarity:	58.14%
Best Local Similarity:	56.89%
Query Match:	97.41%
Idels:	10
Gaps:	0
Matches:	8
Conservative:	1
Mismatches:	3
Indels:	2
Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-15 (1-837)

100	1	ProValValaAnVal1GIgYlnAsnLeuValValAspLeuSerThglnLlePheGlyHis	20
101	11		
102	76	CCCCGGCGAATGtGGGGCAAAACCTGGCTGCTGATCTTTGACGCAAAATCTTTGGCAT	133
103	21	AsnAspTyPProGluThrIleThrAspTyValThrLeuGlnArgG1SerAlaGly	40
104	136	AACCATTAATCCGGAAACCATTAACAGACTATGTCACACTGCACAGAGGCTGGCTTATGGC	193
105	41	GlyValLeuSerAsnAspPheSerGlyThrValLysTySerGlySerSerTyPProPhePro	60
106	196	GGCGTGTTATCTAATTTTTCGGAGCCGAGAAATATAGCGCAGTACGATTCATTTCCCT	255
107	61	ThrThrSerGluThrProArgValValTyPAsnSerArgThrAspLysProThrProVal	80
108	256	ACCACAGCGCAACGCCGCCGCTGTTTAAATTCGAGAACCGATTAACCCGTGGCCGGTG	315
109	81	AlaLeuTyPLeuThrProValSerSerAlaGlyGlyLeuValLleLysAlaG1SerLeu	100
110	316	GCCTGTTAATTTGACGCCCTGTGACACAGTCGGCGGGGCGCATTAAGCTGGCTCATTTA	375
111	101	IleAlaValLeuLleLeuThrArgGlnThrAsnAsnTyPAsnSerAspAspPheGlnPheVal	120
112	376	ATTGCCCTGCTTTATTTTGGCACAGACCAACAACCTTAACACAGGAATTTCCACTTTGTG	435
113	121	TrpAsnLleTyPAlaAsnAsnAspValVal1ProThrGlyGlyCysAspValSerAla	140
114	436	TGCAATATTTTACGCCCAATATATATGtGGtGGtGCTACTGGCGGCTGGAGTGTTCtGGCT	495

[illegible]

RESULT 13
HS-09-000

```

? Sequence 5, Application US/09900575
? Patent No. US20020150587A1
? GENERAL INFORMATION:
? APPLICANT: Langertmann, Solomon
? APPLICANT: Revel, Andrew
? APPLICANT: Auguste, Christine
? APPLICANT: Burtelin, Jeanne
? TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
? FILE REFERENCE: 469201-549
? CURRENT APPLICATION NUMBER: US/09/900,575
? CURRENT FILING DATE: 2001-07-06
? PRIOR APPLICATION NUMBER: US/60/216,750
? PRIOR FILING DATE: 2000-07-07
? NUMBER OF SEQ ID NOS: 64
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 5
? LENGTH: 840
? TYPE: DNA
? ORGANISM: E. coli
? CS-09-900-575-5

```

Alignment Scores:	
Pred. No.:	1 35e-101
Score:	825 00
Percent Similarity:	97.52%
Best Local Similarity:	96.89%
Query Match:	97.29%
DB:	10
Gaps:	0
Indels:	0
Mismatches:	4
Conservative:	156
Matches:	840
Length:	840

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-5 (1-840)

Protein	Accession	Length (aa)	Start	End	Score	E-value	Ident	Mismatches	Gaps	Positives	Negatives
1	ProValValaAnValAlGIgYInaLeuValAAspLeuSerThnGlnIlePheCysHis	20	1	20	100	0.00	20	0	0	20	0
76	CTCGCGGTGAATGATGGGGCAAAACCTGCTGATCTTTCGACCAATCTTTGGCCAT	135	1	135	100	0.00	135	0	0	135	0
21	AsnAspIyrProGluThrIleThrAspIyrValThrLeuGlnArgIleSerAlaIyrGly	40	1	40	100	0.00	40	0	0	40	0
136	AACAGATTATCGGAAACCATATACAGACTATGTCACACTGCACACGAGGCTCGGCTTATGGC	195	1	195	100	0.00	195	0	0	195	0
41	GlyValLeuSerAsnPheSerGlyThrValIyrIyrSerGlySerSerIyrProPhePro	60	1	60	100	0.00	60	0	0	60	0
196	GGCGTGTATCTAATTTTTCGGGACCGGTAATATAGGCGAGTAGTATTCATTCATTCGC	255	1	255	100	0.00	255	0	0	255	0
61	ThrThrSerGluThrProArgValIyrIyrAsnSerArgThrAspIyrProIrpProVal	80	1	80	100	0.00	80	0	0	80	0
256	ACTACCAAGCAAAACGGCGCGGCTGTTTATATTCGAGACAGCATTAACCCCTGCCGCTG	315	1	315	100	0.00	315	0	0	315	0
81	AlaLeuIyrIleuThrProValSerSerAlaGlyIleuValIleIysAlaGlySerLeu	100	1	100	100	0.00	100	0	0	100	0
316	GCCTTTATTTGACGGCTGTGACCACTGGCGTGGGCGGATTAAGAAGCTGCTCATTTA	375	1	375	100	0.00	375	0	0	375	0
101	IleAlaValLeuIleLeuArgIlnThrAsnAnIyrIyrAsnSerAspAspPheGlnPheVal	120	1	120	100	0.00	120	0	0	120	0
376	ATTGCCGTGCTTATTTTGGGACAGAACCAACACTTAACACCGATGATTTCCAACTTTGTG	435	1	435	100	0.00	435	0	0	435	0
121	TrpAsnIleIyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla	140	1	140	100	0.00	140	0	0	140	0
436	TGGAAATTTTACGCCCAATATGATGTGGTGCGCTACTGCGGCTCGCATGTTTCGCT	495	1	495	100	0.00	495	0	0	495	0
141	ArgAspValThrValThrLeuProAspIyrIyrGlySerValProIleIrpLeuThrVal	160	1	160	100	0.00	160	0	0	160	0
496	CATGATGTACACGTTACTCTGCCGACATACCTGGTGTACAGGCCAAATTCCTCTTACCGTT	555	1	555	100	0.00	555	0	0	555	0

QY 161 Tyr 161
Db 556 TAT 558

RESULT 14
US-09-900-575-20
; Sequence 20, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burtlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-20

Alignment Scores:
Pred. No.: 1,836-101 Length: 837
Score: 824.00 Matches: 156
Percent Similarity: 97.52% Conservative: 1
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 97.17% Indels: 0
Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-20 (1-837)

QY 1 ProvalValaSnValaIglYnAsnLeuValaValaSpLeuSerThrgInIlePheCysHis 20
Db 76 CCGCGCGTGAATGTGGGGCAAAACCTGCTGCGTAGATCTTTCAGACGCAAACTCTTTGGCCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValaThrLeuGlnArgGlySerAlaTyrGly 40
Db 136 AACGATTATCCGGAACCATTAACAGACTATGTACACTGCACAGACGCGCGGTATATGGC 195
QY 41 GlyValIleuSerAsnPheserGlyThrValaLysTyrSerGlySerTyrProPhePro 60
Db 196 GCGGTGTATCTAATTTTCGGGACCGTAAATATAGTGGCAGTAGCTATCCATTTCCG 255
QY 61 ThrThrSerGluThrProArgValaValaTyrAsnSerArthrAspLysProTTPProVal 80
Db 256 ACCACCGAGCAAAACGCGCGGTGTTTATTAATTCAGAACGGAATTAAGCCGTGGCGGTG 315
QY 81 AlaLeuTyrLeuThrProValaSerSerAlaGlyLeuValaIleLysAlaGlySerLeu 100
Db 316 GCGCTTATTTCAGCGCTGTGAGCAGTCCGCGCGGTGGCATTAACCTGGCTCATTA 375
QY 101 IleAlaValaLeuIleLeuArgGlnThrAsnTyrAsnSerAspPheGlnPheVal 120
Db 376 ATTCGCCGTGCTTAATTTGCGACAGACCAAAACATATAACGCGATTTCCAGTTTGTG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValaValaProThrGlyGlyCysAspValaSerAla 140
Db 436 TGGATATTATTCAGCCCAATATGATGTGTGTGCTACTGCGCGCTCGATGTTTCGCT 495
QY 141 ArgAspValaThrValaThrLeuProAspTyrArgGlySerValaProIleProLeuThrVal 160
Db 496 CGTGATGTACCGTTACTCTGCGGACTACCGTGTTCAGTGCACAAATTCCTTACCGTT 555
QY 161 Tyr 161
Db 556 TAT 558

RESULT 15
US-09-900-575-16
; Sequence 16, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burtlein, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-16

Alignment Scores:
Pred. No.: 3,4e-101 Length: 837
Score: 822.00 Matches: 155
Percent Similarity: 98.14% Conservative: 3
Best Local Similarity: 96.27% Mismatches: 3
Query Match: 96.93% Indels: 0
Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-16 (1-837)

QY 1 ProvalValaSnValaIglYnAsnLeuValaValaSpLeuSerThrgInIlePheCysHis 20
Db 76 CCGCGCGTGAATGTGGGGCAAAACCTGCTGCGTAGATCTTTCAGACGCAAACTCTTTGGCCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValaThrLeuGlnArgGlySerAlaTyrGly 40
Db 136 AACGATTATCCGGAACCATTAACAGACTATGTACACTGCACAGACGCGGTATATGGC 195
QY 41 GlyValIleuSerAsnPheserGlyThrValaLysTyrSerGlySerTyrProPhePro 60
Db 196 GCGGTGTATCTAATTTTCGGGACCGTAAATATAGTGGCAGTAGCTATCCATTTCCG 255
QY 61 ThrThrSerGluThrProArgValaValaTyrAsnSerArthrAspLysProTTPProVal 80
Db 256 ACCACCGAGCAAAACGCGCGGTGTTTATTAATTCAGAACGGAATTAAGCCGTGGCGGTG 315
QY 81 AlaLeuTyrLeuThrProValaSerSerAlaGlyLeuValaIleLysAlaGlySerLeu 100
Db 316 GCGCTTATTTCAGCGCTGTGAGCAGTCCGCGCGGTGGCATTAACCTGGCTCATTA 375
QY 101 IleAlaValaLeuIleLeuArgGlnThrAsnTyrAsnSerAspPheGlnPheVal 120
Db 376 ATGCTGTCTTAATTTTGGACAGACCAATTAACAGAGATTAAGCTGCTAGTTGTG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValaValaProThrGlyGlyCysAspValaSerAla 140
Db 436 TGGATATTATTCAGCCCAATATGATGTGTGTGCTACTGCGCGCTCGATGTTTCGCT 495
QY 141 ArgAspValaThrValaThrLeuProAspTyrArgGlySerValaProIleProLeuThrVal 160
Db 496 CGTGATGTACCGTTACTCTGCGGACTACCGTGTTCAGTGCACAAATTCCTTACCGTT 555
QY 161 Tyr 161
Db 556 TAT 558

RESULT 16
US-09-900-575-18
; Sequence 18, Application US/09900575
; Patent No. US20020150587A1

GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burtlein, Jeanne
; TITLE OF INVENTION: Fimh Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 18
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-18

Alignment Scores:
Pred. No.: 3.4e-101 Length: 837
Score: 822.00 Matches: 155
Percent Similarity: 98.14% Conservative: 3
Best Local Similarity: 96.27% Mismatches: 0
Query Match: 96.93% Indels: 0
DB: 10 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-18 (1-837)

QY 1 ProvalValAsnValGlyGlnAsnLeuValAlaSplLeuSerThrGlnIlePheCysHis 20
DB 76 CCTCCGCTGAATGGGGCAAAACCTGCTGATCTTTCGACGCCAACTTTGGCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 136 AAGCATTAACCGGAACCATTAAGACTATGTACACTGCACAGAGAGTTGGCTTATGCC 195
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
DB 196 GGGCTGTATCTAATGTTTTCCGGACCGTAATAATATAGTCAGTATCCTTCCCT 255
QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
DB 256 ACTACGAGCAAAACCGCGGCTGTTTAATTCAGAACGATTAAGCCGTGGCGGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
DB 316 GCGCTTATTTGACCGCTGTGACAGCAGTGGGGGAGTAAAGCTGGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnTyrAsnSerAspPheGlnPheVal 120
DB 376 ATGGCGCTGCTTATTTGGCAGACAGCAACACTATTAACAGCATATTCAGTTTGTG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 436 TGGAAATATTACCAATATATGATGTGGTCCACAGCGCGCTGCATGTTTGGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 496 CGTATGTACACCGTACTCTGCGGACTACCGCTTCCAGTCCGATTCCTTACCGTT 555
QY 161 Tyr 161
DB 556 TAT 558

RESULT 17

US-09-900-575-8
; Sequence 8, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine

APPLICANT: Burtlein, Jeanne
; TITLE OF INVENTION: Fimh Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 8
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-8

Alignment Scores:
Pred. No.: 6.31e-101 Length: 837
Score: 820.00 Matches: 155
Percent Similarity: 96.89% Conservative: 1
Best Local Similarity: 96.27% Mismatches: 5
Query Match: 96.70% Indels: 0
DB: 10 Gaps: 0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-8 (1-837)

QY 1 ProvalValAsnValGlyGlnAsnLeuValAlaSplLeuSerThrGlnIlePheCysHis 20
DB 76 CCTCCGCTGAATGGGGCAAAACCTGCTGATCTTTCGACGCCAACTTTGGCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 136 AAGCATTAACCGGAACCATTAAGACTATGTACACTGCACAGAGAGTTGGCTTATGCC 195
QY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerTyrProPhePro 60
DB 196 GGGCTGTATCTAATGTTTTCCGGACCGTAATAATATAGTCAGTATCCTTCCCT 255
QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
DB 256 ACTACGAGCAAAACCGCGGCTGTTTAATTCAGAACGATTAAGCCGTGGCGGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
DB 316 GCGCTTATTTGACCGCTGTGACAGCAGTGGGGGAGTAAAGCTGGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnTyrAsnSerAspPheGlnPheVal 120
DB 376 ATGGCGCTGCTTATTTGGCAGACAGCAACACTATTAACAGCATATTCAGTTTGTG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 436 TGGAAATATTACCAATATATGATGTGGTCCACAGCGCGCTGCATGTTTGGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 496 CATGATGTACACCGTACTCTGCGGACTACCGCTTCCAGTCCGATTCCTTACCGTT 555
QY 161 Tyr 161
DB 556 TAT 558

RESULT 18

US-09-900-575-11
; Sequence 11, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burtlein, Jeanne
; TITLE OF INVENTION: Fimh Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575


```
; ORGANISM: E. coli
US-09-900-575-4
Alignment Scores:
Pred. No.: 8 64e-101
Score: 819.00
Percent Similarity: 98.14%
Best Local Similarity: 95.65%
Query Match: 96.58%
DB: 10
Matches: 840
Conservative: 154
Mismatch: 4
Indels: 3
Gaps: 0
US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-4 (1-840)
QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
Db CCTGCCGATGATGCGGCGCAAAACCGTGTGATCTTTCACCCCAATCTTTGCCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValIleThrLeuGlnArgGlySerAlaTyrGly 40
Db AACGATTACCCAGAAACCATTAAGACTATGTACACACTGCACAGAGTGGCGCTTATGCG 195
QY 41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
Db GCGGTATCTATCTAGTTTTCGCGACCGTAATATATGCGAGTATCTTCCCT 255
QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
Db ACTACACCGCAAAACCGCGGCTGTATTAATTCAGAACGAGTAACCGCGCGGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
Db GCGCTTATTGACCCCGGTGAGCAGTGCAGGAGGAGGCGGATTAAGCTGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
Db ATTGCCGTGCTTATTTCGACAGACCACTATTAACACGATTCATTCAGTTTGTG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
Db TGGAAATATTACCCCAATATGATGTGTGTCGCCACTGCGGCTGATGTTCTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
Db CGTGATGTCCACGTTACTCTGCCGACACTGCCGTTCAAGTCCGATTCCTTACCGTT 555
QY 161 Tyr 161
Db 556 TAT 558
RESULT 21
US-09-900-575-9
; Sequence 9, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burielin, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-9
Alignment Scores:
```

```
Pred. No.: 1.17e-100
Score: 818.00
Percent Similarity: 97.52%
Best Local Similarity: 95.65%
Query Match: 96.46%
DB: 10
Matches: 837
Conservative: 154
Mismatch: 3
Indels: 4
Gaps: 0
US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-9 (1-837)
QY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
Db CCTGCCGATGATGCGGCGCAAAACCGTGTGATCTTTCACCCCAATCTTTGCCAT 135
QY 21 AsnAspTyrProGluThrIleThrAspTyrValIleThrLeuGlnArgGlySerAlaTyrGly 40
Db AACGATTACCCAGAAACCATTAAGACTATGTACACACTGCACAGAGTGGCGCTTATGCG 195
QY 41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
Db GCGGTATCTATCTAGTTTTCGCGACCGTAATATATGCGAGTATCTTCCCT 255
QY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTyrProVal 80
Db ACTACACCGCAAAACCGCGGCTGTATTAATTCAGAACGAGTAACCGCGCGGTG 315
QY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyLeuValIleLysAlaGlySerLeu 100
Db GCGCTTATTGACCCCGGTGAGCAGTGCAGGAGGAGTGCAGTAAAGCTGCTCATTA 375
QY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
Db ATTGCCGTGCTTATTTCGACAGACCACTATTAACACGATTCATTCAGTTTGTG 435
QY 121 TrpAsnIleTyrAlaAsnAsnAspValValProThrGlyGlyCysAspValSerAla 140
Db TGGAAATATTACCCCAATATGATGTGTGTCGCCACTGCGGCTGATGTTCTGCT 495
QY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
Db CGTGATGTCCACGTTACTTTCGCGACTCCCGTTCAGTGCAGTTCCTTACCGTT 555
QY 161 Tyr 161
Db 556 TAT 558
RESULT 22
US-09-900-575-1
; Sequence 1, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burielin, Jeanne
; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-1
Alignment Scores:
Pred. No.: 1.6e-100
Score: 817.00
Percent Similarity: 96.89%
Best Local Similarity: 96.27%
Length: 837
Matches: 155
Conservative: 1
Mismatch: 5
```

Query Match:	96.34%	Indels:	0
DB:	10	Gaps:	0

Oy		1	ProValValAsnValGlyIglAsnLeuValValAspLeuSerThrGlnIlePheCysHis	20
Dd		76	CCCGTCGGATGTGGGGCAAAACCTGTCGTGAATCTTCCAGCCGAATCTTTTGCCAT	135
Oy		21	AsnAspIyrProGluTrnIleIrhAspTryValThrlengInahrgLysAlatyrgLy	40
Dd		136	AACGGTTATCGGAACCATTAACACTATGTCACTGCACGAGGCTCGGCTTATGCC	195
Oy		41	GlyValleuSerAsnPheSergLYThVallysTyrseryLyseryrProphePro	60
Dd		196	GCGCGTTATATTATTTTCCGGATCGTAAATATAGTGGCAGTAGTACTCTTCCCT	255
Oy		61	ThrTrSerGIurhnProxgValValTyAsnSerArghThAspIsProTrproVal	80
Dd		256	AACCACACGAAAGCCGCCGCTGTATTATATTCGAACGAGATAAGCCGTGGCCGGTG	315
Oy		81	AlaleuTYrLeuthProvalSerSerAlagLyIyeuValIleylsAaglYserIeu	100
Dd		316	GCGCTTATTTTGACGCCCTGTAGCAGTGGCGGGGAGTGGCGATTAAGCAGGCTATTA	375
Oy		101	IleAlaValleuIleLeuArgLnThAsnastrYsrAsnsrpaspPhecinPheVal	120
Dd		376	ATTTGCCGGTGCTAATTTTCGCACAGCACACACATATAACGCGATGGTTCCAGTTTGIG	435
Oy		121	TriPasnIleYrAlaAsnaAsnAspValValAlProthrGelYelcYsaSpValSerAla	140
Dd		436	TGGATATTATACGCCAATAATGATGTGGTGGCCCACTGGCGGCTGGCATGCTTGCT	495
Oy		141	ArgasPvalThValThLeuProAspIyrArgelyserValProIleProleuTrnVal	160
Dd		496	C GGATGCACCGTACTCTGCCGAGACTACCTGGTTCAGTGGCGGATCTTACCGITT	555
Oy		161	Tyr 161	
Dd		556	TAT 558	

RESULT 23
US-09-900-575-13
Sequence 13, Application US/09900575
Patent No. US20020150587A1
GENERAL INFORMATION:
APPLICANT: Langemann, Solomon
APPLICANT: Revel, Andrew
APPLICANT: Auguste, Christine
APPLICANT: Burlein, Jeanne
TITLE OF INVENTION: F1mH Adhesin Proteins and Methods of Use
FILE REFERENCE: 469201-549
CURRENT APPLICATION NUMBER: US/09/900.575
CURRENT FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: US/60/216.750
PRIOR FILING DATE: 2000-07-07
NUMBER OF SEQ ID NOS: 64
SOFTWARE: PatentIn version 3.0
SEQ ID NO: 12

```

; seq id no 13
; Length: 837
;
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-13

Alignment Scores:
Pred. No.:      2.18e-100      Length:      837
Score:          816.00        Matches:      154
Percent Similarity: 97.52%      Conservative: 3
Best Local Similarity: 95.65%      Mismatches:  4
Query Match:     96.23%      Indels:      0
DB:              10          Gaps:        0

```

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-13 (1-837)

QY	1	ProValValAsnValGlyIlnAsnLeuValAlaSpIeuSerThrGlnIlePheCysHis	20
QY	111		
Db	76	CGTGGCGTGAATGTGGGGCAAAACCTGGTCGTAGATVCTTTGGACGAAATCTTTGGCCAT	135
QY	21	AsnAspTyrProGluThrIleThrAspIyrValThrIleGlnArgGlySerAlaTyrGly	40
Db	136	AAAGATTACCCAGAAACCACTTACAGACTATGTACACACTGCACACAGAGGTGGGCTTATAGGC	195
QY	41	GlyValIleuSerAsnPheSerGlyThrValIleTyrTyrSerGlySerTyrProPhePro	60
Db	196	AGCGTGTATCTAGTTCTTTCCGGAGCCGTAATAATATATATGACAGAGTATCTTTCCCT	255
QY	61	ThrThrSerGluThrProArgValValTyrAsnSerAArgThrAspLysProThrProVal	80
Db	256	ACTACCCAGCAAAACGCCGGCGGTGTTTATATTTGGAACAGCGAATAACCTCGCCGGTG	315
QY	81	AlaLeuTyrIleuThrProValSerSerAlaGlyGlyLeuValIleLeuAlaGlySerLeu	100
Db	316	GCGGTTTATTTTGACGCCGGCGGAGACAGTGGGGGGAGTGGCGATTAAAGCTGGCTCATTA	375
QY	101	IleAlaValIleuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal	120
Db	376	ATTTCGGCTGTATTGTCGACAGACCAACCACTAATACAGCGATGATTTCCAGTTGGTG	435
QY	121	TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla	140
Db	436	TGCAGATTTTACGCCAATATGATGTGGTGGGCCACACGGCGGTGTGATGTGTTCTGCT	495
QY	141	ArgAspAlaThrValThrIleuProAspTyrArgGlySerValProIleProIleuThrVal	160
Db	496	CGTATGTACACGTTACTCTGCCGGACTACCTCGGTTCAGTGCAGATTCCTTTAACGTT	555
QY	161	Tyr	161
Db	556	TAT	558

```

RESULT 24
US-09-900-575-14
; Sequence 14, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burleau, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; NUMBER OF SEQ. ID NOS: 64
; SOFTWARE: Patentin version 3.0
; SEQ. ID NO. 14
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-14

```

[illegible]

Db 76 CCTGCCGTAATGTGGGGCAAACCTGCTGTAGATCTTTGACGCAAACTCTTGGCAAT 135

```
OY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 136 AAGGATTACCGAAGAACCTTACAGACTATGTACACACTGCACAGAGTTCCGGCTTATGGC 195
OY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
DB 196 ACCGTTATCTAGTTTTCGGGACCGGTAAATATATGACAGTACCTTCCCT 255
OY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTTPProVal 80
DB 256 ACTACACAGCAAGACCGCGGCTTTATTAATTCAGACAGGATTAACCCGGCGCGGTG 315
OY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
DB 316 GCGCTTTATTTGACCGCGGTGACAGTCCGGGAGATGGCGGATTAAGCTGGCTCATTA 375
OY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
DB 376 ATTGCCGCTTATTTTTCGCACAGCAACAATAACACGATTCGATTCAGTTTGTG 435
OY 121 TrpAsnIleTyrAlaAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 436 TGGATATTATTCGCAATATGATGTGGTGGTCCGCTGCTGATGTTTCTGCT 495
OY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 496 CGTGATGTACCGTTACTCTCCGAGCTACCTGTTCCAGTCCGATTCCTTACCGCTT 555
OY 161 Tyr 161
DB 556 TAT 558
```

```
RESULT 25
US-09-900-575-2
: Sequence 2, Application US/09900575
: Patent No. US20020150587A1
: GENERAL INFORMATION:
: APPLICANT: Langermann, Solomon
: APPLICANT: Revel, Andrew
: APPLICANT: Auguste, Christine
: APPLICANT: Buriel, Jeanne
: TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
: FILE REFERENCE: 469201-549
: CURRENT APPLICATION NUMBER: US/09/900,575
: CURRENT FILING DATE: 2001-07-06
: PRIOR APPLICATION NUMBER: US/60/216,750
: NUMBER OF SEQ ID NOS: 64
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 2
: LENGTH: 837
: TYPE: DNA
: ORGANISM: E. coli
US-09-900-575-2
```

```
Alignment Scores:
Pred. No.: 7.5e-100 Length: 837
Score: 812.00 Matches: 153
Percent Similarity: 96.89% Conservative: 3
Best Local Similarity: 95.03% Mismatches: 5
Query Match: 95.75% Indels: 0
DB: 10 Gaps: 0
```

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-2 (1-837)

```
OY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 76 CTGCGGTGAATGTGGGCAAAACCTGCTGATCTTTCACGCAAAATCTTTGGCAT 135
OY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 136 AAGGATTACCGAAGAACCTTACAGACTATGTACACACTGCACAGAGTTCCGGCTTATGGC 195
```

```
OY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
DB 196 GCGCTTATCTAGTTTTCGGGACCGGTAAATATATGACAGTACCTTCCCT 255
OY 61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTTPProVal 80
DB 256 ACTACACAGCAAGACCGCGGCTTTATTAATTCAGACAGGATTAACCCGGCGCGGTG 315
OY 81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
DB 316 GCGCTTTATTTGACCGCGGTGACAGTCCGGGAGATGGCGGATTAAGCTGGCTCATTA 375
OY 101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspPheGlnPheVal 120
DB 376 ATTGCCGCTTATTTTTCGCACAGCAACAATAACACGATTCGATTCAGTTTGTG 435
OY 121 TrpAsnIleTyrAlaAsnAspValValProThrGlyGlyCysAspValSerAla 140
DB 436 TGGATATTATTCGCAATATGATGTGGTGGTCCGCTGCTGATGTTTCTGCT 495
OY 141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
DB 496 CGTGATGTACCGTTACTCTCCGAGCTACCTGTTCCAGTCCGATTCCTTACCGCTT 555
OY 161 Tyr 161
DB 556 TAT 558
```

```
RESULT 26
US-09-900-575-19
: Sequence 19, Application US/09900575
: Patent No. US20020150587A1
: GENERAL INFORMATION:
: APPLICANT: Langermann, Solomon
: APPLICANT: Revel, Andrew
: APPLICANT: Auguste, Christine
: APPLICANT: Buriel, Jeanne
: TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
: FILE REFERENCE: 469201-549
: CURRENT APPLICATION NUMBER: US/09/900,575
: CURRENT FILING DATE: 2001-07-06
: PRIOR APPLICATION NUMBER: US/60/216,750
: NUMBER OF SEQ ID NOS: 64
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 19
: LENGTH: 837
: TYPE: DNA
: ORGANISM: E. coli
US-09-900-575-19
```

```
Alignment Scores:
Pred. No.: 7.5e-100 Length: 837
Score: 812.00 Matches: 154
Percent Similarity: 97.52% Conservative: 3
Best Local Similarity: 95.65% Mismatches: 4
Query Match: 95.75% Indels: 0
DB: 10 Gaps: 0
```

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-19 (1-837)

```
OY 1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
DB 76 CTGCGGTGAATGTGGGCAAAACCTGCTGATCTTTCACGCAAAATCTTTGGCAT 135
OY 21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
DB 136 AAGGATTACCGAAGAACCTTACAGACTATGTACACACTGCACAGAGTTCCGGCTTATGGC 195
OY 41 GlyValLeuSerAsnPheserGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
DB 196 GCGCTTATCTAGTTTTCGGGACCGGTAAATATATGACAGTACCTTCCCT 255
```


Db 544 -----TTGACCAATACAGAGACATATCCATGTCGACGATGATCCAAACAT 597
Qy 112 TyranSerAspAspPheGlnPheValTrpAsnIleTyr-----AlaAsnAsp 128
Db 598 GCAATTTTGCAACATA---TACATTTGGGGGATTCACACCCGACGACGACCAAGAA 654
Qy 139 -----ValValValProThrGlyGlyCysAspValSerAlaArgAspValThr 144
Db 655 CAACACGAGCTGATGTCACACATCAGGAGAGACGACAGCTCTTACCAGAGAACCCAG 714
Qy 145 ValThrLeuProAspPtyrArgGlySerValPro 155
Db 715 CAACATATATATCCGATATCGGGTCCAGACC 747
RESULT 28
US-09-745-763-145/c
; Sequence 145, Application US/09745763
; Patent No. US20020065394A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
McCoy, John M.
Lavallee, Edward R.
Collins-Racle, Lisa A.
Evans, Cheryl
Werberg, David
Treacy, Maurice
Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
ENCODING THEM
NUMBER OF SEQUENCES: 219
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/745,763
FILING DATE: 18-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 145:
SEQUENCE CHARACTERISTICS:
LENGTH: 1904 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 145:
US-09-745-763-145
Alignment Scores:
Pred. No.: 1.98 Length: 1904
Score: 76.00 Matches: 40
Percent Similarity: 39.10% Conservative: 21
Best Local Similarity: 25.64% Mismatches: 44
Query Match: 8.96% Indels: 51
DB: 10 Gaps: 7
US-09-900-575-29_COPY_26_186 (1-161) x US-09-745-763-145 (1-1904)
Qy 37 SerAlaTyrglyGlyValLeuSerAsnPheSerGlyThrVal-----Lys 51

Db 722 TCTGTAAAGGTGGGAAATGAGAGATCTACAGATGAGGTGGAGGAGATACTGCCCT 663
Qy 52 TytSerGlySerSerTytrProPheProThrThrSerGlu----- 64
Db 662 CACTCAGGAACAGCTTATAGACTACAAAGCCGATCCCAAGAGTACCACGATGTAATCA 603
Qy 65 -----ThrProArgValValTyranSerArgThrAspLysPro 77
Db 602 ATCCACTCATGTTACAGAGATCCGCGAGAGACACTTATATATATATATACAGAAAGAG 543
Qy 78 TrpProValAlaLeuTytrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAla 97
Db 542 CAAGCCGCTCTCTTCAGACTCTTCAGTTCTTCAGTTCGACGAGCCAGCTCTGTAATCTA 483
Qy 98 Gly-----SerLeuIleAlaValLeuLeuLeuArgGlnThrAsnAsn 111
Db 482 AATTATATCTCCAGCCCAAGAACCTCTTATGTAATACATGCTGCTTACAGAGACTCTTAC 423
Qy 112 TyranSerAspAsp-----PheValTrpAsnIleTytrAlaAsnAspVal 129
Db 422 CTTCACAGCTCACCACAGATTTTCCAAATTTGATGCAATATC----- 380
Qy 129 lValValProThrGlyGlyCysAspValSerAlaArgAspValThrValThrLeuProAs 149
Db 379 -----TAACTCCGCTTACATTCCTCACT 357
Qy 149 pTyranGlySerValProIlePro-----LeuThrValTyr 161
Db 356 GTAC-----ATCATACCATCCACACCTTTGTTGACACATGTAT 317
RESULT 29
US-10-036-342-79/c
; Sequence 79, Application US/10036342
; Patent No. US20020090681A1
GENERAL INFORMATION:
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Stewart, Timothy A.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3030R1C5
CURRENT APPLICATION NUMBER: US/10/036,342
CURRENT FILING DATE: 2001-12-26
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/112514
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: 60/113300
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/113430
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113605
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113621
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/114140
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/115552
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/116843
PRIOR FILING DATE: 1999-01-22
PRIOR APPLICATION NUMBER: 60/125774
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/125778
PRIOR FILING DATE: 1999-03-23

APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE OF INVENTION: ACTS ENCODING THE SAME
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/052,586
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063564
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063734
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063870
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066120
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066466
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066772
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069335
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069425
PRIOR FILING DATE: 1997-12-12
PRIOR APPLICATION NUMBER: 60/069870
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/068017
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081049

PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086023
PRIOR FILING DATE: 1998-05-18
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086486
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087098
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087208
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05

1	PRIOR APPLICATION NUMBER: 60/0888212
2	PRIOR FILING DATE: 1998-06-05
3	PRIOR APPLICATION NUMBER: 60/0888217
4	PRIOR FILING DATE: 1998-06-05
5	PRIOR APPLICATION NUMBER: 60/0883326
6	PRIOR FILING DATE: 1998-06-04
7	PRIOR APPLICATION NUMBER: 60/0886555
8	PRIOR FILING DATE: 1998-06-09
9	PRIOR APPLICATION NUMBER: 60/0887222
10	PRIOR FILING DATE: 1998-06-10
11	PRIOR APPLICATION NUMBER: 60/088738
12	PRIOR FILING DATE: 1998-06-10
13	PRIOR APPLICATION NUMBER: 60/088740
14	PRIOR FILING DATE: 1998-06-10
15	PRIOR APPLICATION NUMBER: 60/088811
16	PRIOR FILING DATE: 1998-06-10
17	PRIOR APPLICATION NUMBER: 60/088824
18	PRIOR FILING DATE: 1998-06-10
19	PRIOR APPLICATION NUMBER: 60/088825
20	PRIOR FILING DATE: 1998-06-10
21	PRIOR APPLICATION NUMBER: 60/088826
22	PRIOR FILING DATE: 1998-06-10
23	PRIOR APPLICATION NUMBER: 60/088861
24	PRIOR FILING DATE: 1998-06-11
25	PRIOR APPLICATION NUMBER: 60/088863
26	PRIOR FILING DATE: 1998-06-11
27	PRIOR APPLICATION NUMBER: 60/088876
28	PRIOR FILING DATE: 1998-06-11
29	PRIOR APPLICATION NUMBER: 60/089090
30	PRIOR FILING DATE: 1998-06-12
31	PRIOR APPLICATION NUMBER: 60/089105
32	PRIOR FILING DATE: 1998-06-12
33	PRIOR APPLICATION NUMBER: 60/089512
34	PRIOR FILING DATE: 1998-06-16
35	PRIOR APPLICATION NUMBER: 60/089514
36	PRIOR FILING DATE: 1998-06-16
37	PRIOR APPLICATION NUMBER: 60/089538
38	PRIOR FILING DATE: 1998-06-17
39	PRIOR APPLICATION NUMBER: 60/089598
40	PRIOR FILING DATE: 1998-06-17
41	PRIOR APPLICATION NUMBER: 60/089653
42	PRIOR FILING DATE: 1998-06-17
43	PRIOR APPLICATION NUMBER: 60/089908

Alignment Scores:		
Pred. No.:	2.1	1977
Score:	76.00	Matches: 40
Percent Similarity:	39.10%	Conservative: 21
Best Local Similarity:	25.64	Mismatches: 44
Query Match:	8.96	Indels: 5
DB:	12	Gaps: 71

US-09-900-575-29_COPY_26_186 (1-161) x US-10-052-586-503 (1-1977)

```

QY      37 SerialatyrGlyglyValLeuSerAsnPhseercllythral-----Lys 51
           ||| ||||| :::: | |||||
Db      830 TCTGGTAACGGTGGAAATAGAGGAATTACTCAGAGTCGGGTGAGAGGAATFACTGCCCGT 771
           ||| ||||| :::: | |||||
QY      52 TyrSergIySerSetyrProPheprothrhuSerglu-----64
           ::||| ||||| :::: | |||||
Db      770 CACCTCAGGAAACAGCTTATAGACTACAAGAAGCATCCCAAGATACCAGCATGGTATCA 711
           ||| ||||| :::: | |||||
QY      65 -----ThrProArgValAllylrasnserratgthraSpLyPro 77
           ||| ||||| ||||| :::: | |||||
Db      710 ATCCACATCATGTTACAGSAATCCGCCGGAGGCACATTATAATATAATATCACAGAAAGG 651
           ||| ||||| ||||| :::: | |||||
QY      78 TrpProvalAlaleutryleuthrProvalSerSerlaaglyglyleuuAlleyala 97
           ||| :::: | ||||| ||||| |||||
Db      650 CAAGCCGCTGCTTCCACAGACTCTTCACATTCTGTGAGSCCAAGTTCTGTATATCTA 591
           ||| :::: | ||||| ||||| |||||
QY      98 Gly-----SerleullealValleulleuargglnthraSnAsn 111
           ::||| ||||| :::: | |||||
Db      590 AATTATACCCAGGCCAACAGAACTCTTATACATACTGGTCTTCAGAGCACTCAATGC 531

```

QY	112	tyrinsersaspasp-----	phcgin-phievalttrapsnllletrylalaamsnspva	129
		:::		
Db	530	cttcacacctccacacagattttccaattgtatgcacatgc-----		488
QY	129	lvalvalprothrclglylcyaspsvalseralargaspvalthrleuproas		149
			:::	
Db	487	-----	TAAGTCGGTTTACATTCACACT	465
QY	149	ptyrargclyservalprolepro-----	leuthvaltyr	161
			:::	
Db	464	gtac-----atcatatcccatccacagccttggcttgacacatgat		425

RESULT 31

; Sequence 8841, Application US/09815242

; GENERAL INFORMATION:

APPLICANT: Ohlsen, Kari L.

APPLICANT: Wall, Daniel

APPLICANT: Carr, Grant J.

APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Prokaryotes

CURRENT APPLICATION NUMBER: US/09/815,242
FILING DATE: 2001-03-31

PRIOR APPLICATION NUMBER: 60/

PRIOR APPLICATION NUMBER: 60/
PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/
PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/
PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/
PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/
PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/
PRIOR FILING DATE: 2001-02-16

```

NUMBER OF SEQ ID NOS: 14110
;
SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 8841
LENGTH: 1077

```

ORGANISM: *Staphylococcus aureus*

NAME/KEY: CDS

```

; LOCATION: (1)...(1011)
US-09-815-242-8841

```

Alignment Scores:

Preu. No.:
Score:

Percent Similarity

Best Local Similarity

DB: query maccm;

US-09-900-575-29_COPY_26_186 (1-161) x US-09-815-242-8841 (1-1077)

QY 11 ValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr----- 28

Db 580 ATGGAATGATACACGTAATTCAGTAGACATCCTGTAAGTGAATGGTAACAGGAATT 635

QY 29 AsptyrValThrLeuGlnargGlySerAlaTyrGlyGlyValLeuSerAsnPhSerGly 48

Db 640 GATTAGTTAAATTACAATTACAAGTTGCTATGGGTGACGTTACCGTATAACAAGAA 699

Dd	859	ATGGAATTGAAATACAGTATTCAAGTAGAACAATCCTGTACTGATAAATGGTAACAGAATT	918
Oy	29	AspyrValThrLeuglnArglySerAlatyrglyValLeuSerasnPheSergly	48
Dd	919	GATTAGTTAAATTACAATTACAAGTTGCTATGGCGACGTGTACCCTATAAACAA	978
Oy	49	ThrValIysTyISerSerySerTyrProPheProthrTrhSergluthrProAlaval	68
		::: :::: ::	
Dd	979	GATATTAAATTMACAGACACGCATTTGAATTTAGATTAATCTGAAAATCCT	1032
Oy	69	ValTyrSnserArgtyrAspLysProTyrProValAlaleuLeuIleuThrProValser	88
		:: ::	
Dd	1033	---TACAGAACCTTATGCCATCCATCCACGATAAATTGACGAATATCTTGACCA-----	1083
Oy	89	SeralaglyIleuValIlelyala-----	97
		: :: :	
Dd	1084	-----GGTCGATTATGGTGTGCAATAGACACACATTTACTATTATTACGATACCG	1137
Oy	98	-----glySerIleuIleAlaValleuIleuArgIn	108
		:: :: :: :: ::	
Dd	1138	CCATTATTGATTGATGCTAGTCGCAAAATTATCATCATGATA	1179
		RESULT 34	
		US-09-893-238-3/c	
		: Sequence 3, Application US/09893238	
		: Patent No. US20020150973A1	
		: GENERAL INFORMATION:	
		: APPLICANT: Moore, K.	
		: APPLICANT: Nagle, D.	
		: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT AND	
		: TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY	
		: FILE REFERENCE: 7853-237	
		: CURRENT APPLICATION NUMBER: US/09/893,238	
		: CURRENT FILING DATE: 2001-06-27	
		: PRIOR APPLICATION NUMBER: 09/245,041	
		: PRIOR FILING DATE: 1999-02-05	
		: PRIOR APPLICATION NUMBER: 60/093,630	
		: PRIOR FILING DATE: 1998-07-21	
		: PRIOR APPLICATION NUMBER: 60/104,978	
		: PRIOR FILING DATE: 1998-10-20	
		: NUMBER OF SEQ ID NOS: 129	
		: SOFTWARE: FastSeq for Windows Version 3.0	
		: SEQ ID NO 3	
		: LENGTH: 17056	
		: TYPE: DNA	
		: ORGANISM: Mus musculus	
		US-09-893-238-3	
		Alignment Scores:	
		Pred. No.: 101 Length: 17056	
		Score: 74.00 Matches: 30	
		Percent Similarity: 43.56% Conservative: 14	
		Best Local Similarity: 29.70% Mismatches: 38	
		Query Match: 8.73% Indels: 19	
		Gaps: 3	
		DB: 10	
		US-09-900-575-29_COPY_26_186 (1-161) x US-09-893-238-3 (1-17056)	
Oy	2	valValasnaValglcInasleuValValaspleuserThrinIlePhecysHisasn	21
		: :	
Dd	9975	GTCCTCAATCCACACCAAGGCTGATCCCTTGACCTTGAGAATCTTCTTCT	9925
Oy	22	AspTyrProgluThrIlePhraspyr-ValThrLeuglnArglySerAlatyrgly	41
		:::::	
Dd	9924	-----CTTCTCAATATTTTGTGACCTTGAGAGCTTAATAGCTTGCGAGG	9880
Oy	41	yValleuserAsnpheserGlyThrValIysTyISerSerySerTyrProPheProth	61
Dd	9879	AGGAGATFACACGATTGGGGATGCTAGTATGATGATTGTTAACTTATGATGCTT	9820
Oy	61	rThrsErgluthrProArqValValTyrasnserArgrThraspylsProTyrProValAl	81
Dd	9819	GGATTGAGACACCTTAGAGTT-----CAAGATGGTATACCTGGGTATATTGC	9772

```

QY      81 aleuTyLeuthPrOvalSerSeAlaglyGlyleuValilleysalaaqylerleuI 101
Db      9771 T-----ACGTAGCGTGGGGGGCTAGATCTGCATCAATCAATGAAGTTG 9730

QY      101 e 101
Db      9729 A 9729

RESULT 35
US-09-783-066-8
Sequence 8, Application US/09783066
Patent No. US20020142302A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhou, Ping
APPLICANT: Goodrich, Ryle
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyang
APPLICANT: Yang, Yonghong
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Xue, Aidong J.
APPLICANT: Demanac, Radcoje T.
TITLE OR INVENTION: No. US20020142302A1el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 79ICIP2D
CURRENT APPLICATION NUMBER: US/09/783,066
PRIOR FILING DATE: 2001-02-13
APPLICATION NUMBER: No. US20020142302A1 yet Assigned
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 09/552,929
PRIOR FILING DATE: 2000-04-18
NUMBER OF SEQ ID NOS: 9
SOFTWARE: pl_genes Version 2.0
SEQ ID NO 8
LENGTH: 4682
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (310)..(4446)
US-09-783-066-8

Alignment Scores:
Pred. No.:          16.7           Length:         4682
Score:              73.50          Matches:        36
Percent Similarity: 40.76%         Conservative:   28
Best Local Similarity: 22.93%       Mismatches:    66
Query Match:        8.67%           Indels:         27
DB:                  10             Gaps:           8

US-09-900-575-29_COPY_26_166 (1-161) x US-09-783-066-8 (1-4682)

QY      9 leuValValAspleuSerThInlePhecyHisAsnAspyrProgluhrllethr 28
Db      2923 ATTCTGATGGGTTTAGCATATGCTTATTGGCCGAAGAACAAGAGGACTCAGT 2982
QY      29 ASpyTyr---ValThlEuGlARgISerAlatyGlyValleuSerAnPheSer 47
Db      2983 AAATTATGCTTGTTACGTTCCAAGAGAGAT-----GGAAGACTATATGCAATGGAAGC 30366
QY      48 -----glyThValVstYserGilySerSeryrPro-----PhePro 60
Db      3037 CGMCACGGCTTCATCATGCTGTGATCCCAAGCTATCACAGGCTTGCTGATTCCTGGCCA 30966
QY      61 ThlrThSerGIuThPrOrArgValValTyraSerArgrThAsplyrProtPrProval 80
Db      3097 GCCACGAGC-----TTGCCAGTAATAATATGACACAGCTGGCCCAATGACATT 31444
QY      81 AleuTyLr-----leuThPrOvalSerSeAlagly-----Gly 92

```

```
Db 3145 GGAATTTTGGCCGTGAGATGCTGCCACAGCTCCAGGCCAAGGGATMAAACACA 3204
;
Qy 93 LeuValIleuYsAlaGlySerIleuIleAlaValIleuIleuArgInThrAsnTyr 112
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1943
; LENGTH: 19334
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Db 3325 AAC-----ACCATCATGAAATGCTGCTGCATCTGCCTGAT 3360
;
RESULT 36
; US-09-764-869-1945
; Sequence 1945, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1945
; LENGTH: 17397
; TYPE: DNA
; ORGANISM: Homo sapiens
;
US-09-764-869-1945
;
Alignment Scores:
Pred. No.: 121 Length: 17397
Score: 73.50 Matches: 27
Percent Similarity: 45.45% Conservative: 23
Best Local Similarity: 24.55% Mismatches: 45
Query Match: 8.67% Indels: 15
DB: 10 Gaps: 5
;
US-09-900-575-29_COPY_26_186 (1-161) x US-09-764-869-1945 (1-17397)
;
Qy 34 GlnArgGlySerAlaTyrGlyGlyValIleuSerAsnPheserGlyThrValLysTyrSer 53
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1945
; LENGTH: 17397
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Db 4034 AACGAGGGCCACAGAACTGGAGCTCAATACAGAAATTTTCGGGCTCCATTAGGAAGAA 4093
;
Qy 54 GlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsnSerArg 73
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1945
; LENGTH: 17397
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Db 4094 GGAACAGCATGGTCTGTGGCCCCAGAGGCGCAGC---CAGCCTGTGCTGACAGAGGT 4150
;
Qy 74 ThrAspLysProTyrProValAlaIleuTyrLeuThrProValSerSerAlaGlyIleu 93
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1945
; LENGTH: 17397
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Db 4151 TGGTCCCAACCTGGCCCGCAGCAGCTCTTT---CAGCCCTGGCCCAAGTCATGGGCTTT 4207
;
Qy 94 ValIleuYsAlaGlySerIleuIleAlaValIleuIleuArgInThrAsnAsnTyrAsn 113
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1945
; LENGTH: 17397
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Db 4208 -----GCTGTGCTCTCCAAAGCCCTGCTCATTCAGAGGTCATCATCTGC--- 4255
;
Qy 114 SerAspAspPheGlnPheValIleuTyrAsnIleTyrAla-----AsnAsn 127
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1945
; LENGTH: 17397
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Db 4256 -----TTCCCTGCTGCTGAGCTCCACAGCAGCTGCTGCTGACACCGAA 4306
;
Qy 128 AspValValValProThrGlyGlyCysAsp 137
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1945
; LENGTH: 17397
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Db 4307 AGTGTGTGCTGAATGATGCTGGACCGAT 4336
;
RESULT 37
; US-09-764-869-1943
; Sequence 1943, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
```

```
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1943
; LENGTH: 19334
; TYPE: DNA
; ORGANISM: Homo sapiens
;
US-09-764-869-1943
;
Alignment Scores:
Pred. No.: 142 Length: 19334
Score: 73.50 Matches: 27
Percent Similarity: 45.45% Conservative: 23
Best Local Similarity: 24.55% Mismatches: 45
Query Match: 8.67% Indels: 15
DB: 10 Gaps: 5
;
US-09-900-575-29_COPY_26_186 (1-161) x US-09-764-869-1943 (1-19334)
;
Qy 34 GlnArgGlySerAlaTyrGlyGlyValIleuSerAsnPheserGlyThrValLysTyrSer 53
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1943
; LENGTH: 19334
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Db 4034 AACGAGGGCCACAGAACTGGAGCTCAATACAGAAATTTTCGGGCTCCATTAGGAAGAA 4093
;
Qy 54 GlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsnSerArg 73
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1943
; LENGTH: 19334
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Db 4094 GGAACAGCATGGTCTGTGGCCCCAGAGGCGCAGC---CAGCCTGTGCTGACAGAGGT 4150
;
Qy 74 ThrAspLysProTyrProValAlaIleuTyrLeuThrProValSerSerAlaGlyIleu 93
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1943
; LENGTH: 19334
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Db 4151 TGGTCCCAACCTGGCCCGCAGCAGCTCTTT---CAGCCCTGGCCCAAGTCATGGGCTTT 4207
;
Qy 94 ValIleuYsAlaGlySerIleuIleAlaValIleuIleuArgInThrAsnAsnTyrAsn 113
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1943
; LENGTH: 19334
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Db 4208 -----GCTGTGCTCTCCAAAGCCCTGCTCATTCAGAGGTCATCATCTGC--- 4255
;
Qy 114 SerAspAspPheGlnPheValIleuTyrAsnIleTyrAla-----AsnAsn 127
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1943
; LENGTH: 19334
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Db 4256 -----TTCCCTGCTGCTGAGCTCCACAGCAGCTGCTGCTGACACCGAA 4306
;
Qy 128 AspValValValProThrGlyGlyCysAsp 137
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1943
; LENGTH: 19334
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Db 4307 AGTGTGTGCTGAATGATGCTGGACCGAT 4336
;
RESULT 38
; US-09-764-869-1944
; Sequence 1944, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1944
; LENGTH: 19345
; TYPE: DNA
; ORGANISM: Homo sapiens
;
US-09-764-869-1944
;
Alignment Scores:
Pred. No.: 142 Length: 19345
Score: 73.50 Matches: 27
Percent Similarity: 45.45% Conservative: 23
Best Local Similarity: 24.55% Mismatches: 45
Query Match: 8.67% Indels: 15
DB: 10 Gaps: 5
```



```

/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
/ FILE REFERENCE: PC009
/ CURRENT APPLICATION NUMBER: US/09/764,847
/ CURRENT FILING DATE: 2001-01-17
/ Prior application data removed - consult PALM or file wrapper
/ NUMBER OF SEQ ID NOS: 2003
/ SOFTWARE: Patentln Ver. 2.0
/ SEQ ID NO 1281
/ LENGTH: 19882
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-764-847-1281

Alignment Scores:
Pred. No.: 202
Score: 72.50
Percent Similarity: 41.03%
Best Local Similarity: 29.06%
Query Match: 8.55%
DB: 10
Gaps: 6

US-09-900-575-29_COPY_26_186 (1-161) x US-09-764-847-1281 (1-19882)

QY 28 ThrAspTyrValThrLeuGln-----Arg 35
      |||||
Db 12251 ACCGATGAGCGGACACTACAGTCCCGACTAGGTGACGTCCGACGGCGGGCTGTGGCC 12310
      |||||
QY 36 GlySerAlaTyrGlyGlyValLeuSerAsnPheserGlyThrValLysTyr----- 52
      |||||
Db 12311 GGGTCCCGCCATGGGAGGGCTG-----GGGACTATGCAATTCCAAGGCA 12355
      |||||
QY 53 -----SerGlySerSerTyrProphethrProthrThrSerGlu-----ThrProArgValVal 69
      |||||
Db 12356 TGAACGGGGGGA-TCTCGACCTTACCTTGAGAGTCAAGGACAGATCGGCTTCC 12414
      |||||
QY 70 TyrAsnSerArgThrAspLysProThrProValAlaLeuTyrLeuThrProValSerSer 89
      |||||
Db 12415 CATGAGAGCGCGCCCAACACCCCTCCCTGCACCC-----AGCCCTCACACCCAC 12465
      |||||
QY 90 AlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuLeuArgGlnThr 109
      |||||
Db 12466 CCTCACACACTGGCTCCAGCTGCCAGCTCCGCCGACCCCTTGTGTCAAAATGACAA 12525
      |||||
QY 110 AsnAsnTyrAsnSer-----AspAspPheGlnPheValTyrPasnIleTyr 124
      |||||
Db 12526 AAGCTCATGAGAGTTTTAAGCTGATTTTCTGCACAAAGTGGCAGGTTTAT 12576
      |||||

RESULT 43
US-09-925-299-149
/ Sequence 149, Application US/09925299
/ Patent No. US20020055627A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
/ FILE REFERENCE: PA102
/ CURRENT APPLICATION NUMBER: US/09/925, 299
/ CURRENT FILING DATE: 2001-08-10
/ PRIOR APPLICATION NUMBER: PCT/US00/05883
/ PRIOR FILING DATE: 2000-03-08
/ PRIOR APPLICATION NUMBER: 60/124, 270
/ PRIOR FILING DATE: 1999-03-12
/ NUMBER OF SEQ ID NOS: 1556
/ SOFTWARE: Patentln Ver. 2.0
/ SEQ ID NO 149
/ LENGTH: 882
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (14)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc_feature
/ LOCATION: (25)

```



```
Oy 90 AlAGlYGIYLeuValIleLys-----AlAGlYSer---LeuIleAlaValLeuIleLeu 106
    |||||
    ::::: ||||| ||||| |||||
Db 713 TATGGTGGCAAGAACTAGCTATATACGGGGTTCTCCATTAATGCGCGAATTGTTG 654
Oy 107 ArGInThAsn-----AsnTYrAsnSerAspAspPheGlnPheValTrp 121
    |||
    ::::: ||||| ||||| |||||
Db 653 CGCTCACAAGATCGCGCACTATTGACCGAGCTTCATCCTAGCGATTATCCAAATTCCTGCG 594
Oy 122 AsnIleTYrAlaAsnAspValValValProThrGIYGIYCyAspValSerAlaArg 141
    |||
    ::::: ||||| ||||| |||||
Db 593 AATTAATTTAGCGACGACAAAATGTC-----ACCGTAATAATGTGACAAATGGCTTTCAA 540
Oy 142 AspValIThrValIThrLeuPro-----AspTYrArgGIYSerValProIle 156
    |||
    ::::: ||||| ||||| |||||
Db 539 CAAGTCAAAAGCAACGCTTCGGCCAAAGACGCCGAGCTTAGTACTCATC 489
```

Search completed: November 28, 2002, 21:29:40
Job time : 80 secs

